THE UNIVERSITY OF CHICAGO

NAVIGATING NEGATIVITY IN THE SOCIAL WORLD: AN INVESTIGATION OF (A) RETALIATION EVALUATIONS, (B) MISOGYNY IN COMPARISON TO BLACK RACISM, AND (C) MISOGYNY WITHIN THE CONTEXT OF HOME AND WORK DYNAMICS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

CATHERINE FRANCES STEVENSON

CHICAGO, ILLINOIS

MARCH 2024

For my family and for Q

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

To my advisor, Marc Berman, who helped renew my bruised interest in science by embodying what it should be.

To all of my committee—Marc, YC Leong, Emma Levine, and Lin Bian—for their support and insights.

To my collaborators, who inevitably help me to conceive of a question with a different perspective.

To The Environmental Neuroscience Lab for much help and support.

To Madhan

To my cub, Tasha.

ABSTRACT:

Though generally deriding harm, people are, nonetheless, forced to navigate a social world in which harm inevitably exists but also where inflicting it can be unavoidable. In three chapters, I explore: 1) The social-relational aspects that affect how retaliation is evaluated; 2) how misogyny—the bigotry that subsists in the context of the world's most intertwined groups (men and women)—compares to another form of bigotry, black racism; and 3) the dynamics of how men and women—highly dependent on one another in many circumstances—navigate work and home life in the context of misogyny. By exploring how people evaluate retaliation—or the choice not to retaliate—on behalf of self, friend, or stranger in various circumstances, I found that retaliation was favored when open and visible to offender and onlookers and also commensurate with the original offense and condemned when considered surreptitious and harsh. Perhaps as my most compelling results, the more retaliation was endorsed, the higher people rated retaliation for friends (allies) over self- and third-party retaliation. In other words, when people found overall retaliation to be warranted, they rated acting on behalf of a friend higher than the other identities, suggesting that retaliation may be the purview of allies. Conversely, the more overall retaliation was regarded negatively, the greater people regarded *abstaining* from retaliation for oneself favorably. Such a result suggested that when overall retaliation is condemned, choosing not to retaliate for oneself may be an indication of strength. In Chapter 2, I compared four conditions in which a target, either a white woman, a black man, or a black woman experiences an onslaught of slurs, (along with a control, who was a white man and experiences insults). As predicted, Study 1 demonstrated that people considered 1) racism to

be much worse than misogyny across a range of dependent measures, 2) racism to be diluted by the inclusion of misogyny (black-woman condition), and 3) that misogyny serves as something of a "gateway" to other forms of bigotry. In Study 2, I tweaked the language, adding nuance to the original findings. Finally, in Chapter 3, I explored how manipulating worldly prestige, choice, and child-rearing would interact with gender using scenarios describing couples' work and home-life decisions. I was also interested in how the decisions and positions of a given protagonist's partner would affect evaluations. Overarchingly, I found that women were praised for making non-normative decisions when they were pressured or coerced into them or when they were framed as supporting an ambitious man. Conversely, men were denigrated for taking on child-caring or "lesser" job positions but especially when paired with a partner who had a powerful, lucrative career.

GENERAL INTRODUCTION

Harm is very often associated with negative manifestations of morality—specifically the immoral or amoral side of it. Indeed, while scholars disagree about other components of morality, harm (the avoidance of it) and fairness (the enactment of it) are generally regarded as central to morality. In particular, harm is often invoked in the context of the "dark side" as when people endorse violence in apparent defense of conviction (Workman et al., 2021) or commit any number of bad acts from theft of property to physical assault. Yet, while most people condemn harm—or exhibit harm aversion—in hypothetical and simulated scenarios (e.g., Cushman et al., 2011), research suggests that it is not always perceived to be negative, either in its ancient origins or its modern manifestations (Fiske & Rai, 2014). In fact, Fiske & Rai (2014) argue that people may approve some measure of violence in the regulation of relationships even in first-world countries that no do not support systemic violence like honor killing. For instance, if a man severely insults a female stranger, some might regard a punch in the face as fair retribution for the offender. Though some scholars have questioned the study's assumptions—e.g., whether the participants were truly "naïve" to the parameters of the task— (e.g., Perry, 2013), Milgram famously demonstrated that people will inflict "punishment" on a target in the form of electrical shocks at (apparently) no more than the instigation of the experimenter (Milgram, 1963). More recently research has likewise demonstrated that people are willing to inflict pain on another for personal gain, despite prognostications from other participants that this would not occur (e.g., FeldmanHall et al., 2012). Further, there is the far lesser explored—outside of the economic-game literature—of the effects of *psychological* harm such as using slurs, slandering, or spreading gossip.

Likewise, closely associated with harm is the *threat* of harm (e.g., McCullough et al., 2013). Humans must be quick to identify it, assess it, and sometimes have a ready response that requires little reflection. Then again, there are circumstances that allow for—and demand—more reasoning. Fast heuristics that identify a man with a gun as something to be feared are clearly adaptive, but this same swift processing—often over-taking measured reasoning—has obvious drawbacks that may render people susceptible to stereotyping, for instance. In addition, the calculation of future threat can be a paramount decision, requiring conscious weighing of factors. However, conversely, failing to act in a timely manner can also lead to further harm—say in failing to retaliate for an egregious offense where people would consider retaliation warranted, even necessary in order to demonstrate strength of character.

The current research explores observer responses to the latent and more conscious imposition of harm and threat in three distinct contexts: 1) Retaliation, 2) The imposition of slurs, and 3) The home and work life-choices of a heterosexual, married couple. While we don't always explicitly isolate every particular factor just discussed, these chapters are united in that they explore perceptions of harm and threat from uninvolved observers while varying such considerations as relationships, "otherness," and the automaticity of thought. We also don't directly examine conscious vs. unconscious thought processes, yet all three chapters contain some combination of the two, as does most social decision-making. We thus indirectly explore decisions and reactions within the interplay of latent heuristics and conscious beliefs. Perhaps just as importantly, the chapters all investigate observer reactions, which are both under-explored and must be of great significance because a social action inevitably causes a reaction in another (and most often others, plural), almost by definition.

More specifically, in Chapter 1's five studies, we ask the following primary questions: 1) How does the identity of the Agent affect evaluations of retaliation, 2) In general, when do people regard retaliation positively vs. negatively? 3) How does the identity of the retaliator interact with other contextual factors like the transparency of the actions he takes? 4) Is retaliation a form of assistance? 5) Do people evaluate choosing *to retaliate* differently than *choosing not to:* rephrased, is abstaining from retaliation considered an active choice or does it simply constitute a lack of action? In Chapter 2's two studies, we provide minimal information about what most would consider a bad act—using slurs— asking, 1) Is misogyny considered less threatening and more benign that other forms of bigotry, say black racism? 2) How might these differences manifest in day-to-day interaction? 3) Do people even believe that misogyny is a true psychological minority? Finally, Chapter 3 asks 1) What do perceptions of misogyny within families reveal about categorization and power dynamics? 2) How might this translate into life outside the home? Collectively, the chapters provide insight into the perception of harm and threat and how they might serve to affect reputation and relationships with close others and distant strangers.

CHAPTER 1:
STANDING UP FOR ONESELF VERSUS HAVING ANOTHER'S BACK—HOW RELATIONSHIPS AND SITUATIONAL FACTORS AFFECT EVALUATIONS OF RETALIATION

INTRODUCTION

"It was not my intention to do this in front of you. For that I am sorry; but you can take my word for it that your mother had it coming. If you grow up—if you still feel raw about it—I'll be waiting."

— The Bride in *Kill Bill*

For cult fans of *Kill Bill*, among the more anticipated acts of the movie reaches fulfillment when the character, The Bride (actor, Uma Thurman), a one-time assassin, confronts a former colleague (actor, Vivica A. Fox), retaliates for past atrocities by killing her, and then faces her young daughter, who has witnessed the attack. While the line is embedded in black comedy, it captures many of the real facets and subtle calculations involved in a mysterious phenomenon that permeates modern life: retaliation for established wrongs—along with the ubiquitous, nebulous allure of it. Our emotions often call for it. Intellectually, our intuitions frequently sanction and champion it, even as uninvolved witnesses (Gollwitzer & Denzler, 2009). However, as the Bride acknowledges, it is a perilous undertaking. Though she believes the girl's mother "had it coming"—as does the audience in all probability because they share her point of view—she recognizes that the little girl may have a different perspective and seek her own retribution one day (e.g., Stillwell et al., 2008).

Formally, retaliation may be defined as the act of inflicting vengeance—some form of harm by inflicting cost or withholding benefits—on an offender for a perceived transgression on behalf of oneself or another, whether ally or stranger. It is pervasive and primordial, existing across cultures (e.g., Eriksen & Horton, 1993) and throughout modern society as it's found in isolated tribes (e.g. Boehm 1987) and even in other species (Clutton & Parker, 1995; de Waal 1996). Some psychologists argue for very specific definitions of "revenge" that encompass

ultimate or adaptive purposes for it as a construct (e.g., Krasnow et al., 2012; McCullough et al., 2013). The current studies do not directly wade into this debate, so we refer to retaliation or "vengeance" as defined above: a direct response to an offense against oneself or others (e.g., Govier, 2002)

Regardless of its ultimate purpose, the prevalence of this phenomenon is seemingly remarkable, given that retaliation can be costly and dangerous. Those who engage in it risk bodily harm, loss of property, reputational loss, and escalating patterns of intensifying resentment—like feuds— rather than retribution (e.g., Stillwell et al., 2008; Keysar et al., 2008; Elshout et al., 2017) Further, while all social actions exist in a variable environment, the literature suggests that retaliation is peculiarly sensitive to social factors that affect both how it is enacted and how it is evaluated. In one form or another, studies have increasingly addressed perceived valence, pro-sociality, fairness, and authenticity in the context of retaliation and group dynamics in general (e.g., Fiske & Rai, 2011; Baumard et al., 2013; Raihani & BShary, 2015). For one notable example that we've already introduced, it appears from this past research that retaliation and the offense must align in order to escape escalating feuds—feuds that could lead to the involvement of an increasing number of people as, for instance, allies or third parties step in. The age-old eye- for-an-eye can thus be construed as both an endorsement of retaliation and simultaneously an admonition that the severity and nature of retaliation must match and not exceed that of the original offense (e.g., Carlsmith & Darley, 2008). We assert that retaliation must be *warranted*—or more accurately, *perceived* to be warranted. Otherwise, the retaliator dramatically increases the risk of devastating consequences. Further, as the current studies indicate, they simultaneously hazard the censure of observers, who are the very people who stand to confer benefits and administer rewards or punishment based on their perceptions.

Retaliation is ubiquitous

The current research is relevant because retaliation is common. As has been noted throughout recorded history—and this simple example from a film, *Kill Bill*—people evidently equate retaliation with basic, prepotent responses to common life occurrences. Evidence indeed suggests that these tendencies likely developed before the advent of well-established institutions and authorities for arbitrating conflict between and among people. Retaliation on behalf of self and others not only exists but is deeply ingrained in modern psychology, likely having its antecedents in the ancient world (Boehm 1987; Brown 1991). Despite manifest dangers—which can include risks to life and property—widespread studies have made testament to the fact that people do engage in retaliation. They retaliate for themselves (Brown 1991) and for others (Boyd et al. 2003; Sober & Wilson 2008) as both lab and field studies reveal, especially when recourse to authorities is not readily accessible, as stated above.

A broad swath of studies consistently indicates that people—self-retaliators and third-party retaliators alike— are willing to invest their own resources—or engage in what is known as "costly punishment"— to retaliate for perceived lack of cooperation or “unfairness” in the course of play in economic games (e.g., see Camerer 2003). While the prevalence of third-party retaliation or "punishment" in this context varies, seminal studies place it as high as 50-60% of the time (Guth & Tietz, 1990; Fehr & Fischbacher, 2004). Importantly in consideration of the current research, such studies demonstrate that third parties are willing to engage in retaliation and, as noted, *costly* retaliation, which requires them to make some personal sacrifice for their intervention. The frequency with which victims or so-called second parties (direct victims of the offense) retaliate for themselves versus as third parties is less clear, with some arguing both are

relatively common and others not (see Jackson et. al, 2019 for a review). Regardless, there has been much exploration concerning possible motivations**.**

Why Retaliate?

Many scholars believe that such perilous retaliation developed and has endured as a signal and/or message to avoid inflicting harm on the retaliator and/or to deter those who have already wronged the retaliator from repeating the same exploitative behavior against oneself and against others (e.g., Gollwitzer & Denzler, 2009; Shichman & Weiss, 2022). More comprehensively, many conceive of retaliation as a way of restoring justice or balance to relationships (Sell et al. 2009) and also communities that have been disturbed by an original offense between even just two people (Baumard et al., 2013). For instance, one can imagine that the whole community–from tight-knit, ancient village (Rai & Fiske, 2014) to modern office—has an interest in preventing objectionable or norm-violating behavior from installing itself as acceptable. Partially for these reasons, retaliators stand to gain through possible direct and indirect reciprocity (victims or third parties receiving benefits for their retaliatory actions) and long-term reputational gains that confirm or establish them as strong or eligible allies. The literature has suggested that third parties—in particular—stand to profit in all such ways, as they may serve a valuable function in providing *impartial* condemnation or retaliation (e.g., Chen et al. 2020)

Though it is often depricated by lay people and scholars as being an inappropriate or mean-spirited way of responding to wrongdoing, retaliation can be undeniably alluring (e.g., Tripp & Bies, 2009). This may seem a paradox, but our studies suggest that despite revenge's frequently negative associations, uninvolved observers rate retaliation positively or negatively in

response to specific factors. They are capable of condoning it when "warranted" and condemning it when they deem it is not. Further, we found that their judgements of various retaliators—i.e., for self, ally, or third party—varied predictably with respect to one another when retaliation was sanctioned and when it was not. In fact, there is much evidence that retaliation is not intrinsically "right" or "wrong." By this perspective, it is instead the social landscape that establishes if retaliation meets threshold and activates righteous anger or moral outrage (Trivers, 1979) or if it instead it is considered vindictive or selfish, for example. Excessive revenge might prompt derision or even counter-aggression from the injured party (the original aggressor) and disapproval from witnesses.

There have been very few studies that directly compare retaliating for oneself and retaliating as a third party—much less allies. However, the general literature has supplied some relevant findings about how they may be differentially evaluated—a few outlined above. One of the few studies to compare identities, FeldmanHall and colleagues (2014) found that when presented with an array of options including mere compensation to the victim, third parties were more punitive than second-party victims of an economic game. Chen and colleagues (2020) demonstrated that in the context of an economic game people regarded third-party vs. victim (second party) messages to selfish players as more impartial. In turn, third parties were more generous themselves—a kind of norm enforcement. More generally, other studies have widely suggested— though not all, Goeschl, & Jarke (2016)— that third parties are considered trustworthy, impartial, and otherwise pro-social. Once again, this perception might serve to advertise them as desirable allies or community members—or otherwise enhance their reputation (e.g., Jordan et. al, 2016, Raihini and Bshary, 2015; Nelissen, 2008). More immediately, it could increase their likelihood of receiving reciprocal benefits (e.g., Nowak & Sigmund, 2005). Our

strongest hypothesis was, in fact, that retaliation on behalf of others would be judged as consistently more praiseworthy than retaliating on behalf of oneself.

Critically, given this sensitivity and despite a burgeoning canon of research on vengeance as a broad topic, there has been scant evidence on how people differentially evaluate acts of retaliation on behalf of different social categories of people—specifically on behalf of oneself vs. others. If it is, indeed, the social environment that provokes retaliation and how it is received by others, observer judgments should be especially revealing about the nature, purpose, and mechanism of retaliation. Likewise, the *relationship* of the retaliator to the victim should be a paramount consideration. Through a series of five studies, we asked the following questions: 1) How does the identity of the agent affect evaluations of retaliation, 2) In general, when do people regard retaliation positively vs. negatively? 3) How does the identity of the retaliator interact with other contextual factors like the transparency of the actions he takes? 4) Is retaliation a form of assistance? 5) Do people evaluate choosing *to retaliate* differently than *choosing not to:* rephrased, is abstaining from retaliation considered an active choice or does it simply constitute a lack of action?

The Current Research and Hypotheses—

The current studies thus set out to explore the nuances of how uninvolved observers evaluate retaliation within a varying social-relational context: retaliation on behalf of self vs. a friend vs. a stranger. Most other relevant studies have involved economic games while ours consisted of narratives intended to mimic real-life circumstances. In several experiments (Studies 1, 2a, and 2b), we had participants read brief scenarios in which people choose either to retaliate or to *abstain* from retaliation as either the victim, a friend, or a third-party stranger. The

participants then provided ratings of these actions. In Study 1, we explored retaliation in the context of a park theft and the dramatic circumstances involving chasing down and retaliating against the offender. People overwhelmingly endorsed retaliation and condemned failure to retaliate. In Studies 2a and 2b, we explicitly examined the "transparency" or level of direct confrontation in retaliatory evaluations. We speculated that retaliation behind closed doors would be regarded unfavorably. Transparency may be an essential component of proportionality as actions must be visible in order to 1) have a strong impact on bystanders and 2) to undergo evaluation from those who stand to be affected by the retaliation, directly or indirectly. In Study 3, we compared retaliation for others to direct *assistance* of others. Finally in Study 4, we explored how people would differentially evaluate retaliation by various avengers under unfavorable circumstances, (as well as an additional instance of "helping behavior"). In other words, in this final study, we explored patterns under conditions in which the retaliatory act was likely to be regarded as especially blameworthy and inconsistent with the original offense.

In all experiments, our primary measure of interest was a comprehensive evaluation of praiseworthiness. We reasoned that a broad evaluation like "praiseworthiness" would capture participants' overall judgements of retaliating or not retaliating in one of the three forms as self, ally, or third-party stranger. We expected that, regardless of other contributing factors, people would have a more favorable reaction to retaliation when enacted for others. As discussed, evidence suggests that evaluations of retaliation may be ambivalent and complex. It may involve the integration of several factors—including the perceived justice of the act and the severity of the instigating offense. Nonetheless, we reasoned people would still react more favorably to a retaliator potentially putting herself in harm's way on behalf of *another.*

<u>Pro-sociality Hypothesis 1a</u>: People favor retaliation for others over retaliation for oneself

Strengthened by the available evidence in the literature, our first hypothesis was, therefore, that we expected that people would regard retaliating for others—friend or stranger—more positively than retaliating for oneself. All else being equal, we predicted that retaliation for oneself might be construed as more egotistical and selfishly motivated than those of their counterparts. For instance, while people may sympathize and endorse retaliation for oneself as a demonstration of strength and character, retaliation for another is more easily conceptualized as trustworthy or otherwise pro-social as suggested above (e.g., Jordan et al., 2016).

<u>Integrity Hypothesis 1b</u>: Choosing not to retaliate for oneself may be a demonstration of strength

Secondly, we hypothesized that, when provided with the option of *abstaining* from retaliating, the agents—self, friend, or third party—would at least partially present the inverse pattern. We predicted that, in this case, choosing not to retaliate on one's own behalf would be interpreted as more laudable, perhaps because people would consider it self-sacrificing or at least not specifically selfishly motivated. However, beyond simply representing the "opposite" choice from retaliation, we speculated that actively making the deliberate choice *not* to retaliate for *oneself* when there was opportunity could be a particular show of magnanimity. We thought it possible that, under certain circumstances, people might consider it a kind of inhibition, requiring restraint—and possibly demonstrating confidence, honor, and, ironically, strength. In short, under some certain circumstances, consciously deciding *not* to retaliate might indicate that a person has

no need of it and is in a sense "above it." In this way, abstaining from retaliating could be a more powerful message than retaliating for oneself.

Conversely, we predicted that failing to retaliate for someone else (friend or stranger) would be judged less praiseworthy, potentially because it would appear cowardly, anti-social, or unsupportive (e.g., Schlosser & Levy 2016). As the studies progressed, we also honed predictions about how transparency and direct confrontation would affect appraisals. We expected that openly confronting an offender in a measured manner would be regarded favorably. By contrast, we predicted that retaliating inordinately harshly or privately would be considered less admirable. Collectively, we thus refer to these predictions as the "Integrity" Hypothesis. We predicted that it would be relatively more honorable to abstain from retaliation for self and much more dishonorable to fail to retaliate for others.

While these first two hypotheses may present congruent predictions, we predicted that people might evaluate retaliating and not retaliating to some degree independently. As discussed, failing to retaliate did not always prove to be a simple inversion of retaliating.

Hypothesis 2?: Partiality vs. Favoritism?

We initially remained largely uncommitted to any strong a priori hypothesis about how the other two conditions "friend" vs. "third-party" retaliation might compare, theorizing that different psychological inputs might be at work simultaneously. One hypothesis might be that friends would be more closely associated with self-interest and retaliating for them would be construed as favoritism—or if not quite favoritism, obligation to close others. Having no prior investment or personal stake, third parties would be then registered as more pro-social and, their actions evaluated as more praiseworthy. As discussed, third-party retaliation may present an

opportunity to promote one's eligibility as an ally or possibly as a strong and valuable community member or even a dominant one, seeking high status. Particularly if retaliation is primarily a form of "assistance," retaliating for—i.e., helping— a stranger should be regarded as an exemplary exhibition of "goodness" or "strength" that goes beyond the relative obligation of simply helping an ally. In comparison, retaliation for an ally may be regarded as duty and perhaps to be expected and, therefore, less laudable (e.g., Everett et al., 2018; McManus et al., 2020).

Indeed, similar reasoning ironically argues for the opposite position: that retaliation for allies would be most praiseworthy. Theory has long suggested that people have adapted instincts to aid friends and allies through the likelihood that such behavior would maximize inclusive fitness. Inclusive fitness refers to the probability of one's own genes succeeding into the next generation. Perhaps the most obvious example of why aiding close others would be adaptive is they are more likely to share genetic similarity (Hamilton 1964). We thus reasoned that a second possibility would be that retaliation is the particular sphere of allies. If not quite the domain of allies, retaliating for these two categories—ally and third party— may be qualitatively distinct from one another.

In short, retaliation for an ally versus retaliation for a third party could each support different forms of pro-sociality and reputational benefits. Retaliating for an ally might affirm loyalty and service to existing and future friends. It may, in fact, be one mechanism that favors partiality, which is not generally considered good or fair in other contexts (e.g., Shaw & Knobe, 2013). Dangerous and costly as it is, retaliation is something of a valuable resource that perhaps *should* be used sparingly and with *partiality.* Conversely, it's also possible that retaliating for an ally could be hindered by perceptions of favoritism and self-interest. In contrast, retaliating as a

third party might signal impartiality and perhaps potential for being a *future* loyal ally. Both could be construed as a demonstration of domination or leadership, and either might be perceived as selfless. For these reasons, we anticipated that people could rate either one more highly than the other—or even that varying circumstances might favor one or the other.

Overarchingly, our results indicate that retaliation is a delicate mechanism, highly subject to social context. Through five studies, we varied scene, circumstances like transparency, and the nature and severity of retaliation. Collectively, these findings provide support for several conclusions. 1) Retaliation appears to have no static, default valence and can be deemed positive or negative, especially as it is perceived to be "overt" and commensurate with the original offense versus "covert" and inordinately harsh, respectively. 2) Retaliation can seem to be inherently partial in the sense that retaliating for allies was rated most highly and failing to retaliate for allies was rated unfavorably. We speculate that retaliation is something of a particular mandate of allies, imbibing both privilege and duty—and certainly that it differs from third-party retaliation. 3) Further, retaliation is not a variation of "helping behavior;" they function differently. 4) Results also demonstrated that people regarded choosing *not* to retaliate as a deliberate *choice* when the opportunity presented itself. 5) This finding particularly relates to our two most consistent and, perhaps, novel results. The more that people endorsed retaliation, the greater they likewise favored retaliation for friends over self and third party. In contrast, the more they condemned retaliation, the more highly they rated the particular decision *not* to retaliate—particularly on one's own behalf. These results, in turn, suggest that retaliation—often costly—is a valuable resource that should be deployed selectively.

Experimental Studies

STUDY 1: Park Theft

In Study 1, people read about a scenario in which an offender accosts a victim walking in the park and steals his bag. Several people witness the crime—namely, the victim, himself, the victim's friend, and a third-party bystander. Sometime later, the person we will refer to as the "Agent" encounters the thief and is described as having the opportunity to retaliate or not against this perpetrator. We thus varied two factors in each of the Studies 1, 2, and 3: 1) the action (whether the Agent chooses to retaliate vs. to abstain and do nothing) and 2) the identity of the decision-maker, the Agent (self vs. friend vs. third-party bystander) for a total of six conditions in a fully factorial design. Our primary dependent variable of interest was a bipolar rating of praiseworthiness or blameworthiness of the Agent.

We predicted that retaliation would be interpreted as much more praiseworthy when it was enacted on the behalf of others (friend or stranger) than on behalf of the self. Further, we assumed the reverse pattern for the "no retaliation" conditions, presuming that failing to act for others would surely be construed as less supportive than not acting for oneself. As discussed, we also predicted that *not* retaliating for oneself might be more than just inaction and could be a signal or message of strength. Once again, people might believe that not retaliating for the self could indicate that one is has no need to prove oneself though the adoption of such measures. Yet, whether perceived as strong or weak, we expect that these evaluations will be affected by context, like retaliation, in general—for one, how egregious or otherwise provocative the original offense.

We also included a bipolar measure of loyalty/disloyalty. If the second prediction about retaliating for others is true—that is, retaliating for friends is most praiseworthy—we expected that loyalty would demonstrate a similar pattern to praiseworthiness. In this case, we predicted

that it might, in fact, interact with the effect of condition and thus we planned and performed moderation analyses accordingly. While not our primary focus, we also included several other measures to further explore the nuances of participant impressions: bipolar ratings of "morality/ immorality," "heroism/ cowardliness" and "altruism/ selfishness." Once again, we expected that acting for others would be rated higher on these measures. In general, our predictions for these measures were somewhat conjectural, but we anticipated that morality, heroism, and pro-sociality (termed "altruism" to participants) would interact with condition but not necessarily align as tightly with praiseworthiness as loyalty, considering that retaliation seems to integrate many different social factors and as discussed, is not a simple question of "right" or "wrong."

Methods

Participants. 299 participants were recruited from Amazon Mechanical Turk using Turk Prime. The intended number was 300 with an even 50/cell but despite our best effort at randomization and counterbalancing, occasional dropout produced very slight differences in cell sizes (reported below). Participants completed the experiment in a median time of 3 minutes and were compensated $0.25. No participants were excluded from analyses. The sample had a mean age of M=39.2 (SD=37.3) with a minimum requirement of 18 years of age. 143 participants were female.

Design & Procedure.

Each of the first three experiments (Studies 1, 2, and 3) employed a 3 (agent: self vs. friend vs. third party) x 2 (action: retaliation vs. no retaliation) between-subjects design in which each participant was randomly assigned to one of the six resulting conditions, with

approximately 50 participants per cell; as noted, condition sizes differed very slightly as a result of participants occasionally dropping out of the study. We chose Type 3 for regression and ANOVA for theoretical reasons as we were interested in interactions as well as main effects. We used Holm -Bonferroni (Holm, 1979) to correct for multiple comparisons and set the FWE at $p<0.05$. As discussed, all participants read a short, hypothetical vignette about a perpetrator committing a transgression: in each case, the scenario described a random assailant accosting a victim, knocking him to the ground and then making off with his backpack. In Study 1, the vignettes in each of the conditions began with the following set-up:

TABLE 1.1: Initial set-up for each of the six total conditions for Study 1: (Color coded to reflect condition. Agent (retaliator or non-retaliator) appears in red, blue, or green.)

Imagine the following scene: Casey and his friend, Alex, are walking downtown on a street that is crowded with pedestrians. Casey is carrying a backpack. All of a sudden, a man shoves Casey to the ground and steals his backpack. Casey's friend, Alex, and a bystander, Jordan, also clearly see this happen. Before anyone can get their bearings, however, the man who stole Casey's backpack is out of sight.

From here the vignettes diverged according to assigned condition. The first set of brackets applies to the type of "Agent" variable—the language of the "self" condition appearing first, that of the "friend" condition appearing second, and that of the "third-party" appearing third. The differing language used to delineate "action type," is then noted below under "A" and "B." The "retaliation" conditions appeared first and the "no retaliation" conditions appeared second.

TABLE 1.2: Description of Study 1: Conditions as they differed; Agent is color-coded.

Later as [Agent: Casey/Casey's friend, Alex/bystander Jordan] is walking past a food court, he sees the man who stole his backpack sitting at a table with a couple of other people. There is no doubt that this is the same man: the stolen backpack is even sitting next to his chair. [**Agent**] thinks about whether or not to confront the man.

TABLE 1.2 (continued):

A) Retaliation Conditions: **Agent** ultimately decides to shout at him for being a horrible person and demands that he return the backpack he stole.

or

B) No-Retaliation Conditions: **Agent** ultimately decides to let it go and continue walking home.

After reading the scenario, in each of Studies 1-3 participants provided a rating of how praiseworthy or blameworthy they deemed the actions of the respective agent to be as well as how loyal or disloyal. For all participants, the praiseworthiness question appeared first followed by randomized presentations of the four other variables. All questions took the basic format of: "Do you think that [Agent's] actions were [praiseworthy or blameworthy]? In the case of the "loyalty/disloyalty" question, the phrasing was slightly different so that the measure could be applied to circumstances where people acted for themselves (i.e., the self-conditions); The question appeared as follows: "In general life, do you think that [Agent] is a loyal or disloyal friend." Participants registered their answers on 9-point scales, ranging from "Extremely [Blameworthy]" to a midpoint of "Neither [Blameworthy nor Praiseworthy]" to the highest point under the heading of "Extremely [Praiseworthy]."

While our primary DV of interest was praiseworthiness, and secondarily loyalty—especially, as a possible moderator between condition and praiseworthiness—the correlation among the measures was so apparent that we decided to conduct a principal-component analysis. As by far the most dominant, we used Principal Component 1 (PC1) as a DV and supplement to our analyses.

Results

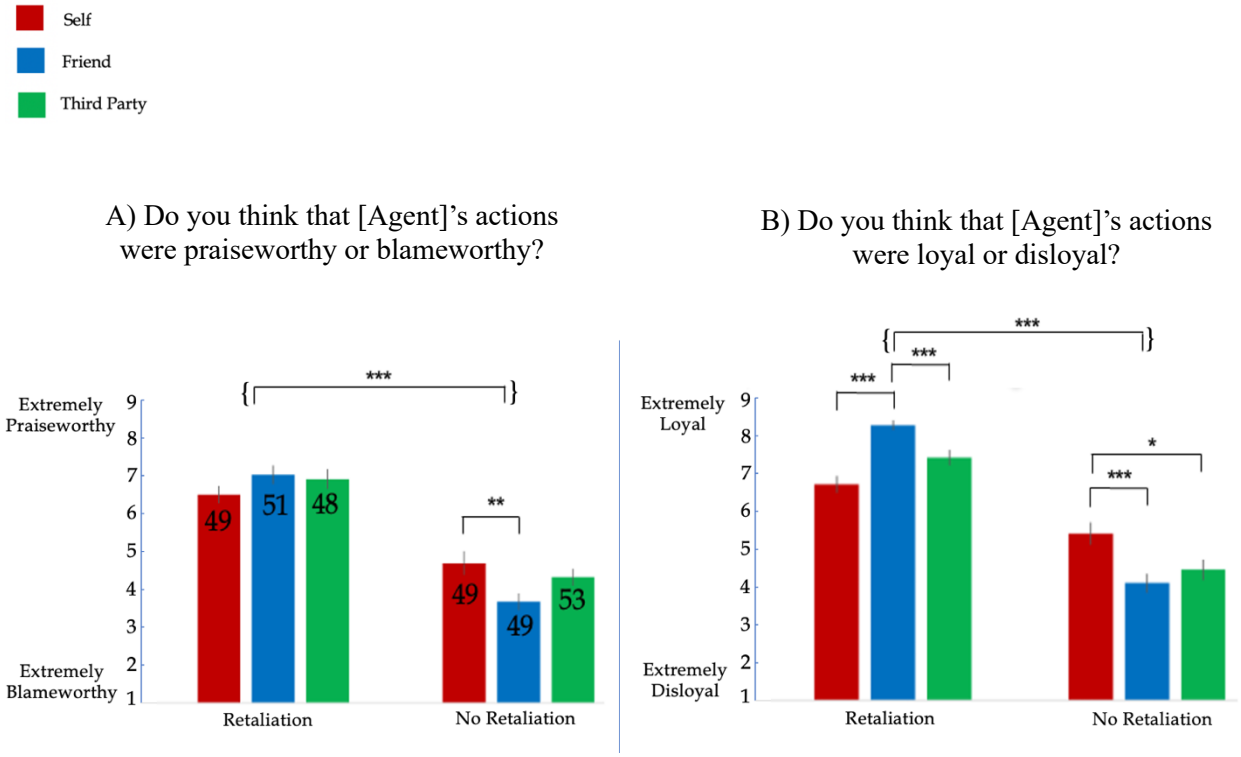

Figure 1.1. Study 1 Evaluations of A) praiseworthiness and B) loyalty as a function of agent identity (self vs. friend vs. third party) and action (retaliation vs. no retaliation); error bars represent *SE's*; numbers represent n's in each respective cell/condition. Additional measures are available in supplemental materials. As printed on Figure1A, among the retaliation conditions, the self condition had an n=49, the friend an n=51, and the third party an n=48. On the non-retaliation side, the self condition had an n=49, the third party an n=49, and the third party an n=53.

Praiseworthiness/Blameworthiness:

The adjusted coefficient of determination for the following model was R^2 =0.361: $F(5, 293)=34.7$, $p<0.001$. Consistent with our hypotheses, a 3(agent: self vs. friend vs. third party) x 2 (action: retaliation vs. no retaliation) factorial ANOVA produced a main effect of action ($F(1, 293)= 26.1$, $p<0.001$, $\eta_p^2=0.081$) and a main effect of agent ($F(1, 293)=4.23$, $p=0.015$, $\eta_p^2=0.028$), as well as an interaction between the two: $F(2, 293)=4.80$, $p=0.009$, η_p^2 =0.032. See Table 3 for these statistics.

TABLE 1.3:

Praiseworthiness model: agent(self vs. friend vs. third party) x action(retaliation vs. no retaliation

PRAISEWORTHINESS *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2=0.361 (adjusted)	$F(5, 293)=34.7$	<0.001	----
Main effect of agent	$F(1, 293)=4.23$	0.015	0.028
Main effect of action	$F(1, 293)= 26.1$	<0.001	0.081
Interaction: agent x action	$F(2, 293)=4.80$	0.009	0.032

Collapsing across the two "action" conditions, overall retaliation (*M*=6.82, *SD*=1.61, 95%CI=[6.56 -7.08]) was deemed much more praiseworthy than choosing not to retaliate (*M*=4.23, *SD*=1.93, 95% CI=[0.308-4.54]; *t*(290)=12.6, *d*=1.46), as suggested by its high effect size. Yet, none of the pair-wise comparisons among the three conditions—self, friend, or third party—were significant. Directionally, these results were arguably consistent with the Pro-sociality Hypothesis as retaliating for self (*M*=6.51, *SD*=2.14, 95%CI=[6.46 -6.97]) was rated less praiseworthy—in magnitude only—than retaliating for either of the other two conditions: third party (*M*=6.92, *SD*=1.5, 95%CI=[6.53-7.31]); *t*(94)=1.31, *p*=0.19, *d*=0.267) and friend (*M*=7.04, *SD*=1.74, 95% CI=[6.60-7.5]; *t*(98)=1.57, *p*=0.1206, *d*=0.077) which was rated highest.

Compellingly and coinciding with our Integrity Hypothesis—the non-retaliation conditions largely mirrored the retaliation conditions in pattern yet produced some significant pairings. As predicted, the agent who chose to abstain from retaliation on his own behalf (*M*=4.70, *SD*=2.19, 95% CI=[4.12-5.27]) was rated as more praiseworthy—or less blameworthy—than either of the other two non-retaliating counterparts in magnitude. Failure to retaliate on one's own behalf was significantly more laudable (or less blameworthy) than failing to retaliate for a friend (*M*=3.67, *SD*=1.56, 95% CI=[3.18-4.16], *t*(98*)*=2.65, *p*=0.009, *d*=0.537). The pair-wise comparison between self and third-party was not significant: *(M*=4.32, *SD*=1.89, 95% CI=[3.92-4.72]; *t*(95*)*=0.98, *p*=0.361, *d*=0.183). Finally, failing to retaliate on behalf a third-party stranger was also marginally less blameworthy than failing to retaliate for a friend (*t*(99*)*=2.65, *p*=0.061, *d*=0.078). Such results were particularly intriguing in that 1)People seem to evaluate retaliation and the decision to abstain from retaliation somewhat independently—at least in this Park scenario; 2) Failing to act for a friend was rated as especially blameworthy; 3) Choosing not to retaliate for oneself was clearly much more praiseworthy than for a friend, as exhibited by a fairly large effect size.

Loyalty

As reflected in Table 1.4A, the following model had an (adjusted) coefficient of determination at R^2= 0.454: $F(5, 293)=50.6$, $p<0.001$. Once again, we ran a 3(agent) x 2(action) factorial ANOVA, yielding a significant main effect of agent $F(1, 293)=8.06$, $p<0.001$, η_p^2 =0.052 and a significant main effect of action, ($F(1, 293)=15.0$, $p<0.001$, η_p^2 =0.049), in addition to a significant interaction between the two: $F(2, 293)=18.3$, $p<0.001$, η_p^2 =0.111. Overall retaliation (*M*=7.48, *SD*=1.48, 95%CI=[7.24 -7.71]), was judged to be far more predictive of

loyalty than non-retaliation (M=4.64, SD=2.00, 95%CI=[4.32 -4.96]): $t(276)=13.9$, $p<0.001$, d=1.61.

TABLE 1.4:

A) Loyalty model: agent(self vs. friend vs. third party) x action(retaliation vs. no retaliation

B) Model exploring loyalty's effect on praiseworthiness as a moderator—i.e., the interaction between loyalty and the six conditions agent x action when regressed on praiseworthiness.

A) LOYALTY *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination= R^2 (adjusted)=0.454	$F(5, 293)=50.6$	<0.001	---
Main effect of agent	$F(1, 293)= 8.06$	<0.001	0.004
Main effect of action	$F(1, 293)= 15.0$	<0.001	0.421
Interaction: agent x action	$F(2, 293)=18.3$	<0.001	0.111
B) MODEL: LOYALTY INCLUDED AS PREDICTOR IN THE FULL-FACTORIAL PRAISEWORTHINESS MODEL			
Effect/Measure	*F*	*p*	η_p^2
Coefficient of determination= R^2=0.573 (adjusted)	$F(11, 287)=37.4$	<0.001	
Main effect of agent	$F(2, 287)=0.950$	0.388	0.007
Main effect of action	$F(1, 287)=6.16$	0.014	0.02
Main effect of loyalty	$F(1, 287)=53.0$	<0.001	0.156

TABLE 1.4B (continued):

Interaction: agent x action	$F(2, 287)=3.67$	0.027	0.025
Interaction: agent x loyalty	$F(2, 287)=1.49$	0.228	0.01
Interaction: action x loyalty	$F(1, 287)=2.66$	0.104	0.009
Interaction: loyalty x agent x action	$F(2, 287)=3.97$	0.020	0.027

People rated the actions of the retaliator acting for himself as least indicative of characteristic loyalty (M=6.71, SD=1.57, 95%CI=[6.27-7.14]) and the friend condition's actions most indicative (M=8.27, SD=1.76, 95%CI=[8.00 -8.51); the two conditions differed significantly at $t(80)=5.97$, $p<0.001$, d=1.21. Participants also considered the actions of the retaliator acting on his friend's behalf to be more diagnostic of general loyalty as compared to the third-party retaliator (M=7.41, SD=1.42, 95%CI=[7.00-7.81]); $t(81)=3.49$, $p<0.001$, d=0.709). In addition, there was a significant difference between the self and third-party conditions: $t(81)=3.29$, $p<0.001$, d=0.468)

The pattern of results for the non-retaliation conditions was essentially an inversion of the retaliation results with the Agent who failed to act for himself considered arguably loyal as the mean was above the midline at M=5.41, SD=2.07, 95%CI=[4.82 -5.60]. The Agent who failed to retaliate on a friend's behalf as most characteristically disloyal (M=4.10, SD=1.75, 95%CI=[3.61-4.59]). This comparison was between the two was significantly different: $t(94)=3.37$, p<0.001, d=0.680). Pair-wise comparisons also revealed that the self condition significantly differed from the third-party condition, (M=4.45, SD=1.97; $t(98)=2.38$, p=0.020,

d=0.472, 95%CI=[3.92-4.98]); the other two conditions did not significantly differ from one another: $t(99)=0.948$, $p=0.345$, $d=0.187$.

Loyalty as an additional predictor in praiseworthiness ANOVA

As stated, we ran a full-factorial ANOVA that included loyalty as a predictor, which yielded a 3-way interaction between loyalty, agent, and action: ($F(2, 287)=3.97$, $p=0.020$, $\eta_p^2=0.027$). (See Table 1.4B). Such a result indicates that the conditions and loyalty act to change their mutual effect on praiseworthiness. Strikingly, as depicted in Table 1.4, there was also a large main effect of loyalty ($F(2, 287)=53.0$, $p<0.001$, $\eta_p^2=0.156$). Notably, including loyalty thus nullified the effect of agent, which was no longer significant. It also reduced the contribution of the other effects from the original model. Furthermore, it reduced the effect of action. Further, adding loyalty increased the fitness of the original model ($R^2=0.361$) which we confirmed by directly comparing them: $F(6, 287)=25.3$, $p<001$. The enhanced model produced an adjusted $R^2=0.573$. Collectively, these results demonstrated that loyalty seemingly featured heavily in participant evaluations of overall praiseworthiness.

Additional Measures: Principal Component Analyses

Altruism, morality, and heroism all correlated closely with praiseworthiness and loyalty (see Figure 1.2A), though, they seemed, once again, to coincide most closely on the non-retaliation side. In terms of retaliation, acting for friend and third party were rated equally altruistic and heroic and significantly higher than acting for oneself. For morality, third-party retaliation was rated higher than retaliation for self, though retaliation for an ally was not. The complete statistics are available in Supplemental Analyses. Here we describe and provide the results for the principal-component analysis (PCA).

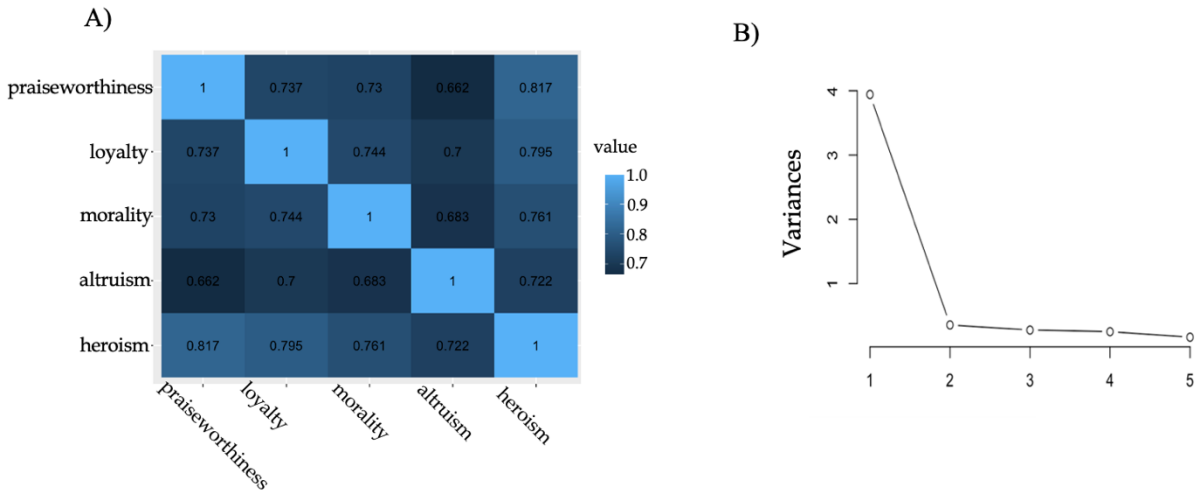

Figure 1.2. A) Correlation matrix for all five measures; B) Scaled scree plot for principal-component analysis (PCA)

We first combined all five measures—praiseworthiness, loyalty, morality, altruism, and heroism—into a PCA that yielded a dominant Principal Component 1 (PC1), accounting for 79% of the variance in the data with a standard deviation of 4.48. PC2 added only 7.00% to the data. Further, the loadings were closely aligned in magnitude and direction, suggesting that PC1 seemed to represent pro-sociality in some form as the measures—with the possible exception of praiseworthiness—could all be construed as specifically relating to socially-beneficial considerations. We, therefore, used only PC1 as a DV to conduct a 3(Agent) x 2(Action) full-factorial ANOVA. The whole PCA was scaled and centered for convenience of interpretation (See Figure 1.3).

Results produced a main effect of agent, ($F(2, 292)=9.54$, $p<0.001$, $\eta_p^2 =0.061$), and a main effect of action, ($F(2, 292)=29.0$, $p<0.001$, $\eta_p^2 =0.090$). The interaction between agent and action was also significant at $F(2, 292)=14.8$, $p<0.001$, $\eta_p^2 = 0.061$. As expected, retaliation ($M=1.33$, $SD=1.26$) and non-retaliation ($M= -1.34$, $SD=1.64$, $t(280)=15.8$, $p<0.001$, $d=1.83$) strongly differed, as supported by a large effect size.

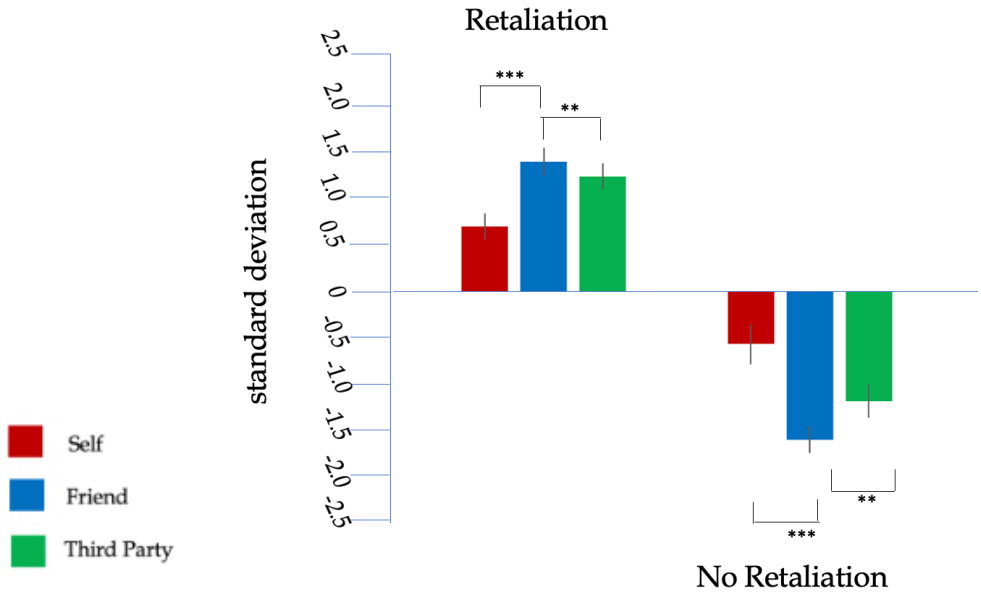

Figure 1.3. Graphical results from full-factorial ANOVA with Principal Component 1 (PC1) as the dependent measure; PC scores were scaled, though when unscaled, analysis provided near identical results. Bars represent *SE's*.

If we accept that PC1 loadings correspond to something like pro-sociality, pair-wise comparisons among all six conditions provide support for the Pro-sociality and Integrity hypotheses. On the retaliation side, acting for oneself (M=0.838, SD=1.18, 95%CI=[0.507-1.17]) was significantly different than both acting for a friend (M=1.67, SD=1.29, 95%CI=[1.32-2.01]); t(97.8)=3.40, p<0.001, d=0.680) and acting for a third-party stranger (M=1.48, SD=1.17, 95%CI=[1.15-1.81]); t(95.0)=2.71, p=0.007, d=0.551). Interestingly, retaliating for a friend and retaliating for a third party did not differ: t(96.9)=2.71, p=0.440, d=0.115.

The non-retaliation side was a near mirror image of the retaliation side with failing to act on one's own behalf, (M=-0.680, SD=1.82, 95%CI=[(-1.18)-(-0.17)]; differing from choosing not to retaliate for a both friend (M=-1.92, SD=1.21, 95%C=[(-2.23)-(-1.60)]; t(83.6)=3.96, p<0.001, d=0.801) and third party (M=-1.42 SD=1.60, 95%CI=[(-1.86 - (-0.98)]); t(96.0)=3.96, p=0.033, d=0.432. The friend and third-party conditions were marginally different: t(95.9)=1.78, p=0.078,

d=0.35. Notably, acting for a friend was greatest in magnitude and lowest in magnitude for retaliation and non-retaliation, respectively.

Study 1 Discussion

Broadly, these results provide some support for our main hypotheses, though the praiseworthiness of retaliation was difficult to interpret. Alone, it did not provide much evidence for the Pro-sociality Hypothesis, but as noted, the dramatic nature of the scenario may account for the lack of differentiation among the retaliators. At this point, we initially speculated that this particular scenario—a theft in a park—may not be a scene that would best exhibit differences among retaliator identities, should they exist. Indeed, this scenario may not necessarily qualify as what most would consider "retaliation." It may instead have been regarded more as justice restoration in the form of retrieving a stolen item under very intense, unusual circumstances. The mere fact that people rated retaliation as so "heroic"—especially for others— suggests that they may have regarded the scenario as an uncommon occurrence and that anyone choosing to chase down a thief is highly commendable. Regardless, if it qualified as retaliation, it was clearly extremely well-received. Further, the PCA results provided support for the Pro-sociality and Integrity Hypotheses, with retaliation for oneself rated significantly less positively than the other two measures and choosing *not* to retaliate for oneself rated more positively than its two counterpart conditions (Figure 1.3).

In accordance with the Integrity Hypothesis, choosing to abstain from retaliation on behalf of the self was evaluated as considerably more acceptable. It was even arguably just short of praiseworthy as with a mean at 4.7, it hovered just over the midpoint on the scale. Failing to retaliate for a friend was rated as most blameworthy followed by third-party lack of retaliation—

though again, the difference between these two was only marginal. Nonetheless, the extreme blameworthiness of not retaliating on behalf of a friend adds intriguing complexity. The results are especially marked and thought-provoking, perhaps especially when considered in the context of the loyalty measure and PCA analysis. The fact that loyalty exerted a strong main effect and interactions with conditions when regressed on praiseworthiness argues for the possibility that retaliation may be the particular domain of allies. More specifically, these findings present the possibility that 1) retaliating for friends could reflect a phenomenon that is specific to allies and that failing to act for them is a severe dereliction of duty, and 2) abstaining from retaliation for oneself might be peculiarly laudable. As another possibility, retaliating for third parties may likewise be a distinct mechanism involving different psychological inputs. While our next studies better address these questions, these results are early evidence for our Hypothesis #2 that people may value partiality in retaliation.

Before speculating further on these findings, we sought to extend and clarify our results in the following studies.

STUDY 2: *Study 2a) Overt, Office-Confrontation Scenarios vs. Study 2b) Covert, "Office-Promotion," Scenarios*

STUDY 2a

Study 2 set out to explore the possible praiseworthiness patterns and to generalize them to another scenario. We were interested in several additional objectives: 1) to examine retaliation more specifically in isolation— in Study 1 we suspected that retaliation, per se, may have been conflated with pure justice restoration as the retaliation conditions included a demand for the stolen backpack; 2) to investigate the implications of using a scenario that is less clearly criminal

(due to the theft) and replace it with a more mundane one; and 3) to examine retaliation in a more naturalistic venue that might more accurately reflect day-to-day interactions in which instances of violation and redress occur more routinely—and potentially, without clear recourse to an authority (e.g., Brezina et al., 2004). Field studies suggest that instances of retaliation increase in the absence of clear protocols or authorities to which victims can appeal (Anderson, 2000). The hypothetical scenario was, therefore, situated in an office setting, where retaliation is known to occur frequently (e.g., Tripp & Bies, 2009). As in the previous study, participants each read a vignette in which the Agent either chose to retaliate or refrain from retaliating for himself, a friend, or a relatively unknown, fellow co-worker.

In addition, we were interested in further investigating the nature of these global "praiseworthiness" and PC1 findings. We speculated that attributions of "strength" or "dominance" may contribute to participant evaluations. We, therefore, substituted past pro-social measures—morality, altruism, and heroism— with two "behavioral" questions (printed below in Table 1.5 below). They were targeted at exploring what kind of behavior people would infer from the differing actions of the agents—specifically, their estimations of the respective agent's characteristic helpfulness and dominance. As earlier discussed, past studies in the literature have suggested that retaliating for oneself and retaliating for another may serve different reputational purposes that may at least partially account for these differences. However, we decided not to include these measures in our primary investigation.

Finally, in conjunction with our other objectives, we investigated the effect of "transparency" by comparing what we are terming, respectively, an "overt" scenario and a "covert" scenario.

Methods

Participants. 297 participants (*M*(age)= 39.0, *SD*=11.9) were recruited through Turk Prime for the Studies 2a and 2b, completing the experiment in a median time of 4.72 minutes. 142 were female. Once again, we had aimed for 300 to produce an even 50/cell, but due to slight drop-out, the cells were not precisely even (see below). Participants were compensated $0.40. None were excluded from analyses.

Study 2a and Study 2b employed the same participants. The two scenarios—overt and covert— were randomized and presented in counterbalanced order, with each having the same six conditions and the same DVs. Participants were randomly assigned to complete scenario 1 or scenario 2 first. Participants read one of the two scenarios first, followed by questions about that scenario, and then went through the same procedure for the second scenario. Finally, they provided demographic information, as in Study 1.

Design & Procedure.

The general procedure for Studies 2a and 2b were identical to that of Study 1, likewise calling for a 3(Agent: self vs. friend vs. third party) x 2(Action: Retaliation vs. No Retaliation) design. The only substantial difference was that participants read about two vignettes in counterbalanced order, with each having the same six conditions and DVs. Participants were randomly assigned to read either the "overt" scenario first—which we are presenting as Study2a—or the "covert" vignette first, which we designate as Study 2b. For each scenario, participants were then randomly assigned to one of the six conditions in the model. Each completed the task of reading a short scenario in which an offender commits a violation against a victim, and the respective Agent chooses to retaliate or not —either on behalf of himself ("self" conditions), on behalf of a friend (the "friend conditions), or on behalf of a co-worker with whom he is relatively unacquainted (the "third-party conditions"). All participants read the same

background information with slight deviations to reflect the respective Agent in play and whether it was a "retaliation" or "no retaliation condition."

After reading each vignette, participants answered six questions, once again using 9-point Likert scales. In this study, we also added additional measures to bolster and generalize the strength of our praiseworthiness and loyalty DVs, respectively. We introduced two additional questions that prompted participants to provide ratings of how "good or bad" the agent's actions were and "how good or bad a friend" they deemed the agent to be. We expected these additions to be highly correlated with our two global measures: praiseworthiness and loyalty, respectively. As reported below, we did find them to be highly correlated. By including these items, we hoped to add credence to our praiseworthiness findings, so they could not easily be attributable to any particular fluke of language we have been using. After completing these steps for the first scenario, they then repeated the same procedure for the other study.

In constructing these two vignettes we were especially interested in manipulating transparency and proportionality between offense and retaliatory response. As one consideration, Study 1 has demonstrated that retaliation, despite its dark connotation, can be highly commendable. As another, it indicated that failing to stand up for a friend might be considered especially reprehensible. We set out to explore these findings by using an incident of retaliation that we expect people to endorse and one that they may regard as less favorable. We, therefore**,** framed Study 2a's "overt" retaliation to be open and specifically calibrated to be a commensurate response to the original wrongdoing. In contrast and as discussed in greater detail in Study 2b, the vignette that we are terming "covert" involved an opportunity for the Agent to provide a negative review to his boss about the offender. Critically, the meeting between Agent and boss occurred in secrecy, added to which the "Adam"—the offender— was not even present. As

stated, these vignettes form the basis of Study 2b below. We theorized that people might disapprove or have misgivings about this "covert" retaliation and regard it with—if not total disapproval—at least more ambivalence and less overall favorability.

We were interested not only in how these differing circumstances in the "overt" scenario would affect judgments of the Agent in various conditions but also in specifically exploring transparency: namely, in what ways the "overt" scenario would systematically differ from the "covert" findings.

The wording for the Study 2a's overt vignette appears in Table 1.5 below.

TABLE 1.5: General template for the vignettes in each of Study 2's "overt" conditions; note: names have been changed in this scenario. Jamie is the "Agent" in each scenario, just as "John" was in the previous scenario.

SELF: Jamie and Nick are long-time co-workers at a company; the two don't know each other very well but have been recently assigned to jointly lead a very lucrative project for an important client. The project involves a large team and much creative and arduous work. Technically having the higher job title, Jamie has the authority to oversee the team and remove people at his discretion. At an early meeting, Nick shocks Jamie by taking credit for a lot of his work, mocking him each time he tries to state an opinion, and basically dictating the next steps that he thinks him should be taking in the project.	*FRIEND*: Jamie and Casey are long-time co-workers at a company, who have become good friends over the years and mutually relied upon one another's help and support. They also work with Nick, though neither of them knows him very well. The three have recently been assigned to jointly lead a very lucrative project for an important client. The project involves a large team and much creative and arduous work. Technically having the highest job title of the three, Jamie has the authority to oversee the team and remove people at this discretion. At an early meeting, Nick shocks Jamie by taking credit for a lot of Casey's work, mocking Casey each time he tries to state an opinion, and	*THIRD PARTY:* Jamie, Casey and Nick are long-time co-workers at a company. The three don't know each other very well but recently have been assigned to jointly lead a group project for an important client. The project involves a large team and much creative and arduous work. Technically having the highest job title of the three, Jamie has the authority to oversee the team and remove people at his discretion. At an early meeting, Nick shocks Jamie by taking credit for a lot of Casey's work, mocking Casey each time he tries to state an opinion, and basically dictating the next steps he thinks that Casey should be taking on the project.

TABLE 1.5 (continued):

At this juncture, each participant was then assigned to one of the following two conditions:

Retaliation condition:

Jamie is mortified by Nick's behavior and knows it to be wrong. In front of the whole group, he sternly reprimands Nick and throws him off the project.

Non-Retaliation condition:

Jamie is mortified by Nick's behavior and knows it to be wrong yet decides that it's not worth confronting Nick. He doesn't say anything in response.

As described above, participants were then prompted to answer questions on 9-point Likert scales. The praiseworthiness and "loyalty," along with the two additional "goodness" and "friendship" questions, were counterbalanced and presented first. Participants then completed the "deterrence" and "pro-social" questions in randomized order. The exact wording for these additional measures appears below in Table 1.6 below.

TABLE 1.6: Additional measures. They read as follows:

Deterrence measure:

"Imagine the following scenario: a co-worker gets annoyed at Jaime during a meeting. Do you think it likely that the co-worker would risk offending Jaime by confronting him?"

Pro-social measure:

"Imagine the following scenario: one of Jaime's co-workers is struggling with a problem at work. Do you think that Jaime would be likely to help out?'

Results

Praiseworthiness/Goodness of Action

Unfortunately, because of a slight typo, the 2 "self" conditions for retaliation and no retaliation were not correctly worded. Although we did not expect that this minor alteration had interfered with participant interpretation of the respective vignettes, we re-ran the self conditions, (pictured in dark red in Figure 6; the original conditions appear alongside in the lighter red). Indeed, the corrected self conditions did not differ from the old, as is readily apparent in Figure 4. Using the "new," self conditions, the overall pattern of results likewise did not change appreciably.

The original wording of the beginning of each flawed self scenario read as follows; the mistake is highlighted in red—"John" should have been "Jamie:"

TABLE 1.7: Set-up for scenario with typo

Jamie and Nick are long-time co-workers at a company; the two don't know each other very well but have been recently assigned to jointly lead a very lucrative project for an important client. The project involves a large team and much creative and arduous work. Technically having the higher job title, Jamie has the authority to oversee the team and remove people at his discretion. At an early meeting, Nick shocks Jamie by taking credit for a lot of his work, mocking him each time he tries to state an opinion, and basically dictating the next steps he believes John should be taking in the project.

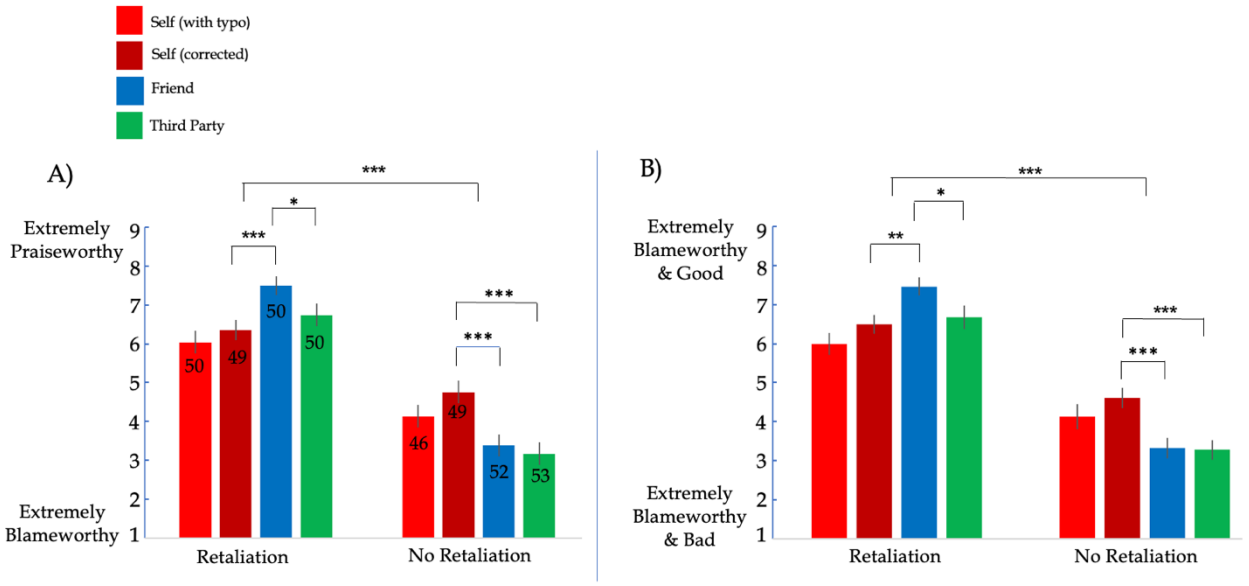

Figure 1.4. Study 2a, Overt Vignette: A)Evaluations of praiseworthiness, alone, and B) Composite evaluations of praiseworthiness & the goodness-of-action measure as a function of agent identity (self vs. friend vs. third party) and action (retaliation vs. no retaliation). Error bars represent *SE's* of the means. As printed on Figure1A, among the retaliation conditions, the self condition had an n=50, the corrected self condition had an n=49, the friend and third party conditions each had an n=50, respectively; on the no retaliation side, the self condition with the typo had an n=46, the corrected self condition had an n=49, the friend condition had an n=52, and the third party condition had an n=53.

The praiseworthiness and goodness-of-action measures were highly correlated at r= 0.880 (p<0.001, 95%CI=[0.862-0.903]), and we, therefore, proceeded to combine the DVs for the remainder of the "positivity" analyses. As depicted in Figure 1.4, praiseworthiness, graphed alone, and praiseworthiness combined with the goodness-of-action measures produced essentially identical results.

Exhibited in Table 1.8A, the adjusted R^2=0.422 for the model was significant at F(5, 295)=50.0 , p<0.001. The 3x2 factorial ANOVA provided a significant main effect of action—F(1, 295)=25.4, p<0.001, η_p^2= 0.079—and a main effect of agent—F(2, 297)=8.39, p<0.001, η_p^2= 0.053. The interaction between the agent and action variables was also significant: F(2, 295)=4.80, p<0.001, η_p^2= 0.062.

TABLE 1.8:

A)Positivity model for overt scenario: agent(self vs. friend vs. third party) x action(retaliation vs. no retaliation)

B)Positivity model with loyalty/friend composite as an added predictor

A) POSITIVITY (Praiseworthiness + Goodness-of-action) *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2=0.442 (adjusted)	$F(5, 297)=50.0$	<0.001	----
Main effect of agent	$F(2, 297)=8.39.$	<0.001	0.053
Main effect of action	$F(1, 293)= 25.4$	<0.00	0.079
Full interaction: agent x action	$F(2, 293)=4.80$	<0.001	0.061
Interaction: agent (self vs. friend) x action	$F(1, 196)=20.4$	<0.001	0.081
Interaction: agent (self vs. third party) x action	$F(1, 197)=7.95$	0.005	0.061
Interaction: agent (friend vs. third party) x action	$F(1, 201)=1.92$	0.167	0.009
B) POSIVITY WITH LOYALTY AS AN ADDED PREDICTOR *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2=0.678 (adjusted)	$F(5, 291)=58.8$	<0.001	----
Main effect of agent	$F(2, 291)=3.78$	0.024	0.003
Main effect of action	$F(1, 291)= 0.336$	0.562	0.143
Main effect of loyalty	$F(1, 291)=7.19$	0.008	0.414

TABLE 1.8B (continued):

Interaction: agent x action	$F(2, 291)=0.068$	0.934	0.022
Interaction: agent x loyalty	$F(1, 291)=5.02$	0.007	0.047
Interaction: agent x action x loyalty	$F(2, 291)=0.252$	0.777	0.001

Collapsing across the two "action" conditions, overall retaliation (*M*=6.88, *SD*=1.95, 95%CI=[6.58-7.18]) was deemed much more praiseworthy than choosing not to retaliate (*M*=3.72, *SD*=1.95, 95%CI=[3.45-4.02]; *t*(290)=12.6, *d*=1.65, as suggested by its high effect size. There was also, of course a main effect of action in the model (above in Table 1.8A). Expanding upon these results, the actions of the retaliator on behalf of a friend (*M*=7.46, *SD*=1.66, 95% CI=[7.13-7.80]) were deemed significantly higher than both than acting for oneself (*M*=6.5, *SD*=1.67 95%CI=[6.03-6.97]; *t*(94)=4.02, *p*<0.001, *d*=0.577) and the third-party condition, *M*=6.67, *SD*=2.16, 95%CI =[6.07-7.26]; *t*(91.9)=2.05, *p*=0.043, *d*=0.411. The third-party retaliator's ratings did not from those of the retaliator who acted for himself, *t*(92.1)= 0.439, *p*=0.661, *d*=0.089, though they just barely trended in the predicted direction.

On the non-retaliation side, failure to retaliate was overall judged to be blameworthy, as we suspected it might, given the characteristics of the this "overt" retaliation. Once again, the actions of the agent who chose to abstain on his own behalf (*M*=4.61, *SD*=1.82, 95%CI=[4.10 - 4.67]) were considered more praiseworthy than his non-retaliating counterpart in the third-party condition (*M*=3.27 *SD*=1.90, 95%CI=[2.76-3.78]; *t*(99.8)=3.63, *p*<0.001, *d*=0.719) and friend condition: (*M*=3.32, *SD*=1.86, 95%CI=[3.06-3.57]; *t*(89.3)=3.50, *p*<0.001, *d*=0.697). Interestingly, the friend and third-party conditions were almost even: *t*(103)=0.145, *p*=0.884—

which was not the case in the park scenario but perhaps was more consistent with the PCA analysis. Otherwise, the two studies presented similar patterns on the non-retaliation side. They likewise produced the same relationships among self, friend, and third party, though, of course, these differences were not significant among the retaliation conditions in Study 1.

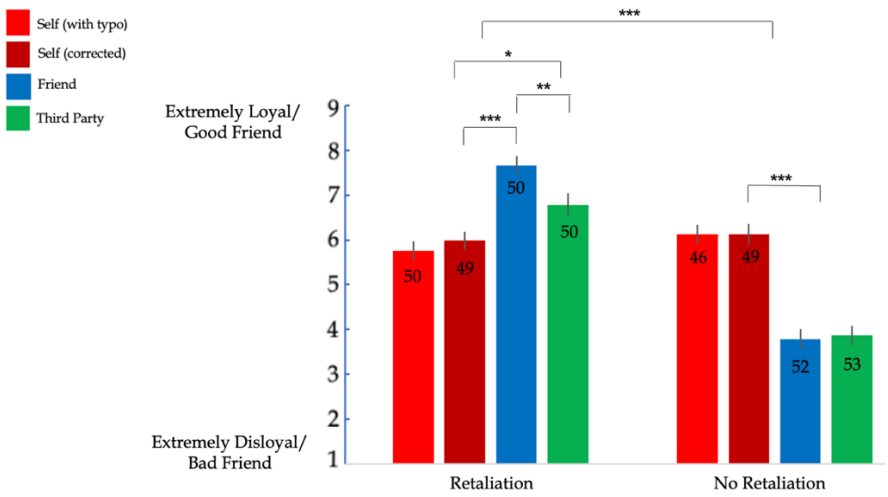

Figure 1.5. Study 2a: Evaluations of loyalty/good-friend combined measures as a function of agent identity (self vs. friend vs. third party) and action (retaliation vs. no retaliation). Bars represent *SE's*. The n's for each cell are, once again, printed on the bar for each condition.

Overt vignette: Loyalty/Likelihood of Being a Good Friend Composite

The "loyalty" and "friend" ratings were likewise highly correlated at Pearson's r= 0.790, p<0.001, 95%CI=[0.744-0.829]; furthermore, the pattern for the two measures was almost exactly the same. Accordingly, they were combined for the following analyses.

The adjusted coefficient of determination for the following model was R^2=0.446: F(2, 297)=49.6, p<0.001. A 3 x 2 factorial ANOVA produced a significant main effect of agent (F(2, 297)= 5.38, p=0.045, η_p^2= 0.035), a main effect of action (F(1, 295)=127, p<0.00, η_p^2= 0.301), as well as a robust interaction between the two: F(2, 295)=42.4, p<0.001, η_p^2=0.223. As exhibited in Table 1.9, overall retaliation (M=6.85, SD=1.41, 95% CI=[6.58-7.13]) was much

greater than overall non-retaliation(M=4.58, SD=1.93, 95%CI=[4.47-4.89]; t(299)=10.7, p<0.001, d=1.24.

TABLE 1.9: Loyalty model: agent(self vs. friend vs. third party) x action(retaliation vs. no retaliation

LOYALTY/FRIEND COMPOSITE MEASURE *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2(adjusted)=0.446	F(5, 297)=43.1	<0.001	----
Main effect of agent	F(2, 297)= 5.38	0.005	0.035
Main effect of action	F(1, 297)=150	<0.001	0.336
Full interaction: agent x action	F(2, 297)=43.1	<0.001	0.225

The retaliation results produced a similar pattern of results to those of the praiseworthiness and positivity, respectively. As in Study 1, they were, once again, aligned with the positivity combined measure, though in this case, both DVs produced significant results. The retaliator acting for a friend (M=7.71, SD=1.41**,** 95% CI=[7.32-8.10]) was evaluated significantly more globally loyal and likely to be a good friend than both the retaliator acting for himself (M=6.01, SD=1.46 95%CI=[5.60-6.42]; t(96.7)=5.88, p<0.001, d=1.18) and the third-party retaliator (M=6.81 SD=1.87, 95%CI=[6.29-7.32]; t(91.06)=2.71, p=0.008, d=0.52). Judging from the effect sizes, retaliation for a friend was especially dominant. The third-party retaliator was also deemed significantly more loyal than the retaliator acting upon his own behalf: t(92.5)=2.36, p=0.020, d=0.474.

Within non-retaliation, the retaliator acting for himself (M=6.15, SD=2.08**,** 95%C1=[5.68 -6.62]) was considered significantly more likely to be loyal than either of the other two conditions, friend (M=3.80, SD=1.65, 95%C1=[3.41-4.19]); $t(98.9)=7.25$, $p<0.001$, d=1.44) and third party (M=3.88 SD=1.87, 95%C1=[3.54-4.22]; $t(99.1)=7.16$, $p<0.001$, d=1.42). As with praiseworthiness, the friend and third-party conditions were almost even. Notably, choosing not to retaliate for oneself was judged well above the midline into loyal, $t(48)=7.20$, $p<0.001$, d=1.30; in addition, it was even rated slightly higher than retaliation for oneself in magnitude, though the two did not significantly differ.

In further analysis, we, once again, ran a full-factorial ANOVA that included loyalty as a possible moderator. The coefficient of determination for the model was R^2=0.678 (adjusted). This model not only produces an interaction with action ($F(2,289)=8.58$, $p<0.001$, $\eta_p^2=0.003$) along with a main effect ($F(1,289)=33.1$ $p<0.001$, $\eta_p^2=0.102$), but it also renders the other effects of action and the interaction between action and agent insignificant. In conjunction with the above analyses, these findings seem to suggest that loyalty continues to figure prominently in participant appraisals of praiseworthiness—or this case, positivity. The full results from the model are in Table 1.8B.

Study 2a: Discussion

In accordance with the Pro-sociality Hypothesis, retaliating for oneself was clearly rated less positively (now by the combined standard of two measures) than acting for another, friend. In addition, as discussed, there were significant interactions between self and third party as well that at least indicate that participants regarded them differently in the context of retaliation vs no retaliation. Once again, failing to retaliate for oneself was significantly less blameworthy than

the other two non-retaliating conditions, satisfying the Integrity Hypothesis. The other two conditions did not differ, a departure from Study 1.

Perhaps even more meaningfully, retaliating for a friend was rated significantly more praiseworthy than either of the two conditions—including third party—which was also true of the PCA results in the Park Study. This result was all the more relevant given that overall retaliation was rated much more highly than non-retaliation. Such findings lend strong credence to the concept that retaliation is a *partial* mechanism and one to be used judiciously. This idea is especially notable because people tend to value *impartiality* under most circumstances. It is not totally clear what may be inducing the exceptionally high rating for acting on behalf of a friend, accompanied by the particularly low rating in the case of *choosing* not to act (especially in Study 1's park scenario). Yet, the relationship between victim and Agent seems to be of undeniable importance. The combined results from the two studies suggest that retaliation may, indeed, be the particular prerogative of allies and may comprise an imperative of action. Regardless, interpersonal relationships appear to contribute significantly to the operation and evaluation of retaliation.

Moreover, the importance of these relationships seems to function very differently for retaliation than it does for helping others. Prior research has found that helping a friend is considered less praiseworthy than helping a stranger because people think that one is obligated to help a friend (e.g., McManus et al., 2020). Contrary to this research, the Agent who retaliated on behalf of a friend was distinctly rated as more praiseworthy than those of any of the other conditions; (We directly compare acts of retaliation and helping in Studies 3 and 4). From this standpoint, it is especially notable that the loyalty ratings very much aligned with the praiseworthiness ratings—as supported by the loyalty analyses and loyalty's interaction when

regressed on praiseworthiness. These loyalty results may provide further support for retaliation as the particular domain of allies—and the corollary possibility that retaliation is inherently a partial mechanism. When approved, people seem to regard retaliating for a friend as a positive form of partiality—loyalty—rather than a negative one—favoritism. In addition, retaliation and non-retaliation do not seem to precisely mirror one another as might be expected, suggesting that people evaluate them slightly differently or that different psychological inputs may be in play. We reserve further discussion of these results for Study 2b—specifically, under conditions in which people were more ambivalent about their overall support of retaliation.

STUDY 2b*: Covert Office Scenario*

As discussed, Study 2b employed the same participants who were used in Study 2. They followed exactly the same procedure as those in Study 2a, except that, once again, some read the overt scenario first and others read this one, the covert scenario, first.

The covert scenario sought to explore the nuances of the most salient results from Studies 1 and 2 and to extend those findings to a new vignette. We were especially interested in whether our most persistent findings from the first two studies—especially as related to the friend condition on the retaliation side and the self-condition on the non-retaliation side—would replicate under circumstances that were specifically *covert* as described in Table 1.10 below. We also theorized that people might rate the overall choice to retaliate and not retaliate differently. More specifically, we predicted that retaliation would be less praiseworthy and abstaining for retaliation would be less blameworthy than in the other two studies. Nonetheless, those findings had contributed to our predictions that people would at least be more ambivalent about retaliation in the covert scenario and might even find abstaining from retaliation for self to be overall more positive than retaliating for self.

We designed the "covert" scenario to occur privately and, critically outside the knowledge not only the offender but also other witnesses. We expected that people would find this lack of openness—particularly concerning the offender suffering retaliation—to be questionable if not offensive behavior. Unlike the scenario in Study 2a, it could be interpreted as surreptitious and incongruent with the instigating transgression. Most prominently, in Studies 1 and 2a, participants evidently considered retaliation in general to be very praiseworthy and failure to retaliate to be blameworthy. We were interested in the patterns that would emerge under circumstances where retaliation is, perhaps, not as clearly commendable (Nelissen, 2008).

TABLE 1.10: Set-up for each of the six total conditions for Study 2b: "Covert Retaliation"

(Note: in this framing, the retaliating or non-retaliating "Agent" is always "John"—in red when he is retaliating for himself; the victim is otherwise always "Mike," appearing in blue when John's friend and green when the "third party")

SELF: John works in the marketing department at Company X. A senior position has recently opened up, and his boss is considering current employees for the promotion. A co-worker, Adam, is one of the two primary candidates under consideration. John doesn't know Adam very well, but they have not generally hit it off when they've encountered one another. At company events and meetings, Adam has had a habit of targeting John for ridicule and generally being unnecessarily rough and difficult with him.	**FRIEND:** John and Mike are close friends who work in the marketing department of the same company X. A senior position has recently opened up, and their boss is considering current employees for the promotion. Adam—one of their co-workers whom neither John nor Mike knows very well—is one of the two candidates up for consideration. While Mike and Adam are only casual acquaintances, they have not generally hit it off when they've encountered one another. At company events and meetings, John has observed that Adam has a habit of targeting Mike for ridicule and generally being unnecessarily rough and difficult with him.	**THIRD PARTY**: John, Mike, and Adam are casual acquaintances who work in the marketing department at Company X. A senior position has recently opened up, and their boss is considering current employees for the promotion. Adam and another co-worker are the two primary candidates under consideration. While Mike and Adam don't know each other very well, they have not generally hit it off when they've encountered one another. Adam and another co-worker are the two primary candidates under consideration. At company events and meetings, John has observed that Adam has a habit of targeting Mike for ridicule and generally being unnecessarily rough and difficult with him.

TABLE 1.10 (continued):

From here, those in the *retaliation* conditions viewed the following:

> **Retaliation Conditions**: While in his boss's office, the subject of the promotion comes up. [Agent: John] believes that Adam and the other candidate—who currently have the same title and responsibilities—have similar qualifications; however, he tells his boss that he thinks that the other candidate is more deserving and that his boss should choose him over Adam for the position.

While participants in the *non-retaliation* conditions read this excerpt:

> **Non-retaliation Conditions:** While in his boss's office, the subject of the promotion comes up. [Agent: John] believes that Adam and the other candidate—who currently have the same title and responsibilities—have similar qualifications and tells his boss so. He says he can't reasonably assert that one deserves the position over the other.

Covert Scenario: Results for Praiseworthiness & Blameworthiness & Loyalty:

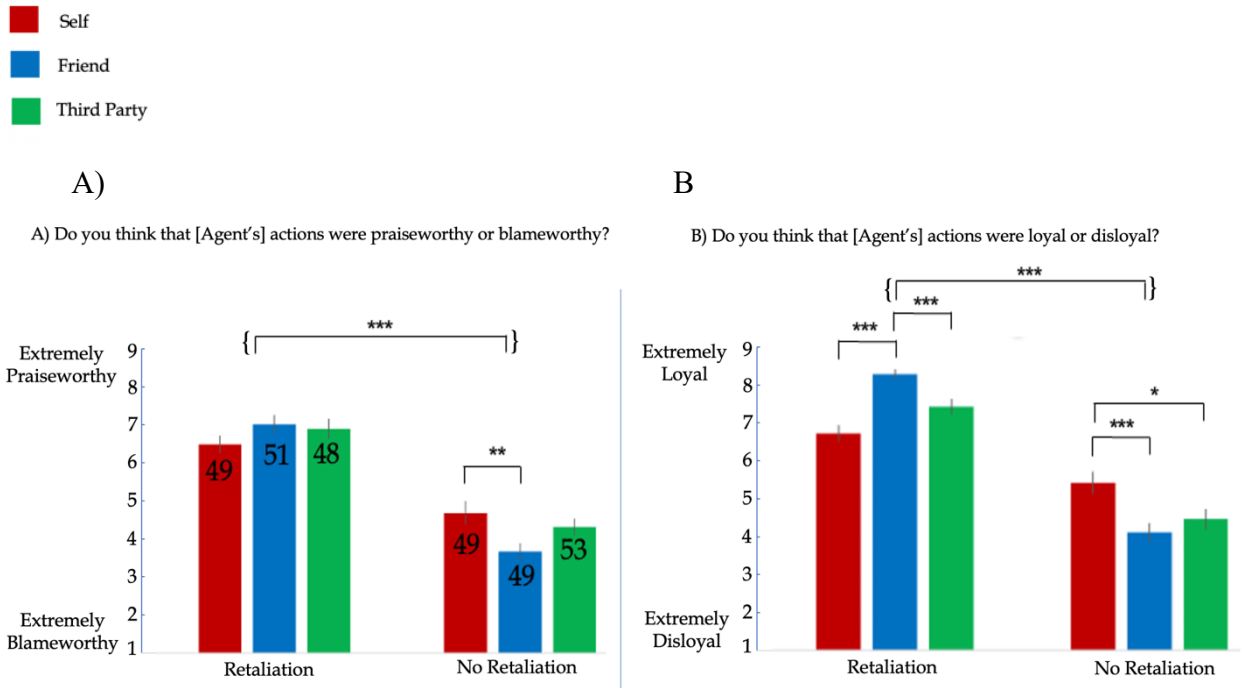

Figure 1.6. Study 2b—Evaluations of combined measures of A) praiseworthiness and goodness of action and B) loyalty and likelihood of being a good friend as a function of agent identity (self vs. friend vs. third party) and action (retaliation vs. no retaliation). Bars represent *SE's* of the means. N's remain the same as those in Study 2a, as once again, participants completed both tasks. The only difference is that, as discussed, we re-ran the "self" conditions in Study 2a.

***Study 2b Results:* Vignette 2— "Office-Promotion", Covert Scenario**

Praiseworthiness & Goodness of Action

Because the additional measure: "Do you think that [Agent's] actions were good or bad" was again found to be highly correlated (r=0.871, p<0.001) with the praiseworthiness measure for the covert vignette, we combined the DVs to form one consolidated measure of positivity.

R^2(adjusted)=0.051 was significant for the following model at $F(5, 295)$=4.211, p<0.001, notably accounting for much less variance than in previous studies. A 3x2 factorial ANOVA provided a significant main effect of agent, ($F(2, 295)$=9.75 , p<0.001, η_p^2= 0.062), a significant main effect of action, ($F(2, 295)$=3.89, p=0.049, η_p^2= 0.013), as well as a significant interaction between agent and action: $F(2, 295)$=5.41, p=0.005, η_p^2= 0.035.

TABLE 1.11:

A)Positivity model for overt vignette: agent(self vs. friend vs. third party) x action(retaliation vs. no retaliation)

B)Enhanced Positivity model with loyalty/friend composite as an added predictor

A) POSITIVITY (Praiseworthiness & Goodness-of-action Combined) *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2=0.678(adjusted)	$F(5, 295)$= 0.051	<0.001	----
Main effect of agent	$F(1, 295)$=9.75	<0.001	0.062
Main effect of action	$F(1, 295)$=3.89	0.049	0.013
Full Interaction: agent x action	$F(2, 295)$=5.41	0.005	0.035

TABLE 1.11A (continued):

Full Interaction: agent x action	$F(2, 295)=5.41$	0.005	0.035
Interaction: agent (self vs. friend) x action	$F(1, 194)=17.2$	<0.001	0.081
Interaction: agent (self vs. third party) x action	$F(1, 194)=6.88$	0.009	0.034
Interaction: agent (friend vs. third party) x action	$F(1, 201)=1.92$	0.167	0.009

B) ENHANCED MODEL: POSIVITY WITH LOYALTY AS AN ADDED PREDICTOR

Effect/Measure	*F*	*p*	η_p^2
Coefficient of determination= R^2=0.450(adjusted)	$F(11, 289)= 23.3$	<0.001	----
Main effect of agent	$F(2, 289)= 0.814$	0.444	0.006
Main effect of action	$F(1, 289)=3.91$	0.05	0.013
Main effect of loyalty	$F(2, 289)=20.7$	<0.001	0.067
Interaction: agent x action	$F(2, 289)=2.56$	0.080	0.017
Interaction: agent x loyalty	$F(2, 289)=1.34$	0.265	0.009
Interaction: action x loyalty	$F(2, 289)=3.40$	0.066	0.012
Interaction: agent x action x loyalty	$F(2, 289)=2.93$	0.055	0.020

As one clear departure from Studies 1 and 2a, both retaliation and non-retaliation—are rated to be praiseworthy, though only narrowly. In fact, a pair-wise comparison reveals that when collapsed across action, retaliation (M=5.65, SD=1.70, 95%CI=[5.37-5.92]) did not differ from the choice not to retaliate (M=5.43, SD=1.93, 95%CI=[4.91 -5.95]; t(279)=1.04, p=0.300, d=0.120), despite the significance of "action" in the model, as appears in Table 1.11A.

While the retaliation pattern was the same in direction as in the previous two studies, none of the comparisons were significant. The retaliator acting for a friend (M=5.94, SD=1.81; 95%CI=[5.42-6.46]) was rated higher than either of the other two conditions, but neither comparison reached significance: neither with self (M=5.46 SD=1.57, 95%CI=[5.02-5.89]); t(91.2)=1.38, p=0.172, d=0.281), nor third-party condition M=5.57 SD=1.73, 95%CI=[5.09-6.05]); t(93.8)=1.01, p=0.311, d=0.207), respectively. In this case, there was, once again, no significant difference between the self and third-party conditions: t(97.1)=0.333, p=0.740, d=0.067.

Yet again, on the non-retaliation end, the Agent who chose to abstain on his own behalf (M=6.15, SD=1.77, 95% CI=[5.71-6.58]) was more praiseworthy than either of the non-retaliating counterparts, third party (M=4.63, SD=1.98, 95% CI=[4.08-5.17]; t(100)=4.13, p<0.001, d=0.814) and friend (M=5.49, SD=1.75, 95% CI=[5.00-5.98]; t(101)=1.90, p=0.060, d=0.375, though this latter comparison did not quite reach significance and was only marginal. Notably, non-retaliation in Study 2b seemed to present a different pattern from the first two studies, with failing to act on behalf of third party rated more blameworthy than failing to act for a friend, (t(98)=2.33, p=0.022, d=0.207)—a kind of "staircase" pattern. Finally, in a distinctive deviation from either of the other two studies, participants rated *abstaining* from retaliation for oneself as significantly more positive than retaliating for oneself: t(101)=2.10, p=0.039,

d=0.412. In fact, in this scenario, choosing not to retaliate was judged to be highest of all six conditions (in magnitude), a provocative result and further support for the for the two original hypotheses and the sensitivity of retaliation in general.

Vignette 2, Covert: Loyalty/Likelihood-of-Being-a-Good-Friend Composite

As expected, the two measures were highly correlated at Pearson's 0.770 (p<0.001), and we, therefore, proceeded to combine the DVs for the remainder of the "loyalty" analyses for the first, "covert" vignette.

R^2(adjusted)=4.17 was significant for the following model at $F(5, 295)=4.21$, $p<0.001$. A 3 x 2 factorial ANOVA produced a significant interaction between the "agent" and "action" variables: $F(2, 295)=7.80$, $p<0.001$, $\eta_p^2= 0.05$, in addition to a main effect of both agent ($F(1, 295)=9.35$, $p<0.001$, $\eta_p^2= 0.060$) and a main effect of action that fell just short of significance: $F(2, 295)=3.76$, $p=0.053$, $\eta_p^2= 0.013$).

TABLE 1.12: Composite Loyalty/Friendship model

LOYALTY/FRIENDSHIP COMPOSITE			
Effect/Measure	*F*	*p*	η_p^2
Coefficient of determination=R^2 (adjusted)=4.17	$F(5, 295)= 4.21$	<0.001	----
Main effect of agent	$F(1, 295)=9.75$	<0.001	0.062
Main effect of action	$F(1, 295)= 3.89$	0.049	0.013
Full interaction: agent x action	$F(2, 295)=7.80$	0.005	0.035
Interaction: agent (self vs. friend) x action	$F(1, 194)=17.2$	<0.001	0.081

TABLE 1.12 (continued):

Interaction: agent (self vs. third party) x action	$F(1, 194)=6.88$	0.009	0.034
Interaction: agent (friend vs. third party) x action	$F(1, 201)=1.92$	0.167	0.009

Once again and as exhibited in Figure 1.6, the loyalty patterns largely aligned with the praiseworthy patterns. Consistent with the previous studies, the retaliator acting on behalf a friend was rated, in terms of direction, as more loyal and likely to be a better friend (M=6.35, SD=2.08 95%CI=[5.88-6.82]) than either the retaliator acting for himself (M=5.69, SD=1.24, 95%CI=[5.34-6.03]); $t(74)=1.88$, p=0.063, d=0.389) and the third-party retaliator (M=5.68 SD=1.69, 95%CI=[5.21-6.15]); $t(172)= 3.67$, p=0.086, d=0.821), yet in both cases the comparisons were marginal. The self and third-party conditions did not differ, ($t(90)=0.033$, p=0.973), being nearly dead even.

As predicted, the Agent choosing not to retaliate for himself was judged to be significantly more prone to loyalty and being a good friend, (M=6.31, SD=1.58, 95%CI=[5.92-6.69]), than either of the other two conditions, friend (M=5.22, SD=1.56, 95%CI=[4.97-5.47]; $t(98.6)=3.70$, p<0.00, d=0.731) and third party (M=5.04, SD=1.67, 95%CI=[4.58-5.50]; $t(98)=4.17$ p<0.001, d=0.821). In this case, the friend and third party did not differ: $t(99.0)=0.728$, p=0.467. Once again in alignment with positivity ratings, abstaining from retaliation for oneself was also judged as a better indicator of loyalty in life than retaliating for oneself: $t(100)=2.37$, p=0.020, d=0.464. Also consistent with the positivity analyses, there was no difference in the ratings between this Agent who abstained from retaliating for himself—though slightly lower in magnitude—and the friend who retaliated: $t(80.0)=0.11$, p=0.913.

Once again, loyalty featured prominently when added as a predictor to the positivity model (See Table 1.11B). There was a main effect of loyalty ($F(1, 289)=20.7$, $p<0.001$, $\eta_p^2= 0.067$. Adding loyalty further rendered the main effect of agent insignificant, though the main effect of action was largely unaffected. Loyalty did not produce any interactions, but in two cases the results were marginal: The full three-way agent x action x loyalty ($F(2, 289)=2.93$, $p=0.055$, $\eta_p^2= 0.019$), and a two-way interaction with action ($F(1, 289)=2.93$, $p=0.055$), $\eta_p^2= 0.079$). There was no interaction between action and agent: $F(2, 289)=2.56$, $p=0.055$), $\eta_p^2= 0.017$.

Further Results

Comparing the effect of transparency ("overt" vs. "covert") circumstances between the "open confrontation" scenario (Study 2a) and the "office promotion" scenario (Study 2b)

SIDE BY SIDE PRAISEWORTHINESS/GOODNESS OF ACTION: For sake of better visualization, below is a side-by-side depiction of the positivity (praiseworthiness/goodness) graphs from the "overt," open confrontation scenario and the "covert," office-promotion vignette, respectively. Once again, we refer to the latter as "covert" because the potential retaliation occurred privately and the former as "overt" because retaliation occurred not only directly in view of the offender but also the victim and other bystanders. Through the course of these studies, we had so far found that people tend to rate retaliation more highly when in view of others—and likely when the retaliatory response was consistent with the original offense. Under these circumstances, people may endorse retaliating for allies and disparage not retaliating partially because visibility determines what is warranted (as a duty)—or merited (as an inclination), depending on perspective. Such results also coincide with "audience effects," in which people tend to increase what they believe to be socially desirable behavior in the presence of witnesses (e.g., Kurzban & Descioli, 2007). We, therefore, might expect that people would be

more likely to construe behavior as objectionable when secretive or "covert." The findings from Study 2a seemed to resemble the results from the Park Study (Study 1) more closely, with choosing not to retaliate overall rated negatively and failing a friend judged very blameworthy. In contrast, the covert vignette produced a stair-case pattern, with choosing not to retaliate for an ally rated somewhere between not retaliating for oneself and not retaliating for a third party. More importantly, both retaliation and no retaliation, when collapsed across action, were likewise rated highly—or at least above the midpoint, which was a major deviation from the previous two studies. We thus set out to directly compare what we are terming a "transparency" factor: "overt" vs "covert" scenarios.

A)

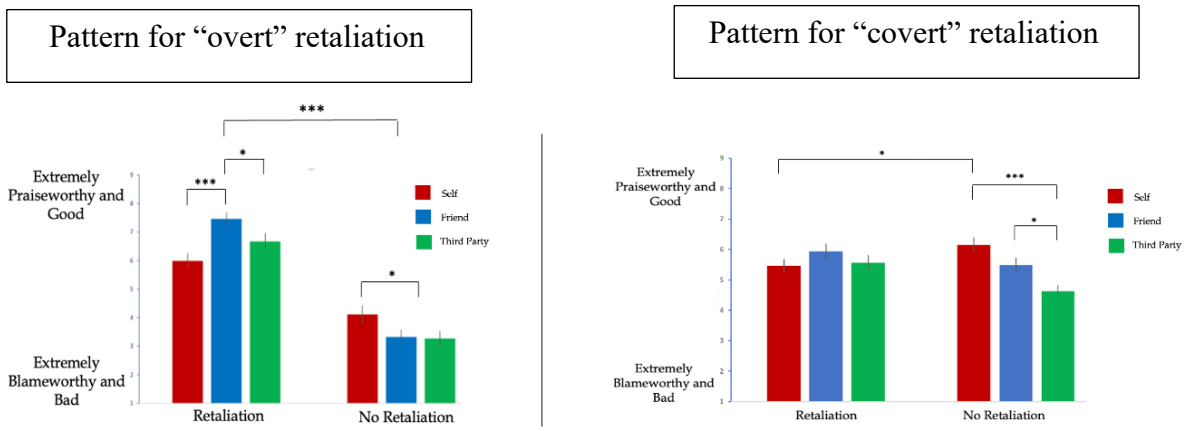

Figure 1.7. Study 2a vs. Study 2b: comparison of praiseworthiness/goodness of action from two vignettes, A) Overt, "Open-confrontation" vignette, and B) Covert, "Office-promotion" vignette, each as a function of agent identity (self vs. friend vs. third party) and action (retaliation vs. no retaliation). Bars represent *SE's* of the means.

Overt vs. Covert Scenarios: Methods

We conducted a mixed-effects regression with transparency (overt vs. covert) as a within-subjects variable and agent (self vs. friend vs. third-party acquaintance) and action (retaliation vs.

no retaliation) as our consistent between-subjects variables. The analysis consisted of a 2 (transparency: overt vs. covert) x 3(agent: self vs. friend vs. third party) x 2(action: retaliation vs. no retaliation) full-factorial fixed ANOVA. We were most interested in how transparency would affect overall evaluations of action, but also how it might affect evaluations of the Agent, particularly the finding that retaliation for a friend seems to be condoned to a greater degree than the other conditions. We also included "order"—as in the order in which participants viewed vignette 1 versus 2—as a covariate. While it was not significant as a main effect ($F(1, 289)=2.50$, $p=0.11$) and did not interact with any of the other variables, we report the results of the enhanced model that controls for it`. One note: we were forced to use results from the self conditions that contained the mistake because there was no way to extricate and replace them with the new ones in the overall mixed design. We were confident with the validity of using these conditions especially because they did not differ significantly from the revised self conditions that we re-ran and used in Study 2a, as earlier tested**.**

Controlling for order, there was no 3-way interaction among action, agent, and transparency: $F(2, 531)=1.39$, $p=0.250$. Of greater interest, results yielded an interaction between transparency and action: $F(1, 515)=96.7$, $p<0.001$, $\eta_p^2 =0.100$. As expected, there was also an interaction between transparency and agent: $F(2, 547)=3.06$, $p=0.048$, $\eta_p^2 =0.017$, as well as an interaction between retaliation and action ($F(3, 541)=8.72$, $p<0.001$, $\eta_p^2 = 0.020$. There was also a main effect of transparency, which emphasizes the stark differences between the two scenarios and how this factor seems to have affected both overall retaliation and the social-relational dynamics we have been exploring: ($F(1, 295)= 7.50$, $p=0.007$, $\eta_p^2 = 0.069$). Collapsed across all six conditions respectively, the overt scenario ($M=5.54$, $SD=1.82$, 95%CI=[5.24-5.65])was rated

overall more highly than the covert scenario, (M=5.12, SD=2.55, 95%CI=[4.83-5.41]; t(543)=2.30, p=0.022, d=0.188).

Expanding upon these interactions, we first investigated differences in retaliation and lack of retaliation between the two scenarios. As discussed and as is readily visible in Figure 7, retaliation vs. choosing not to retaliate does not always produce inverse results; for instance, in the case of this comparison of scenarios, the results for the retaliation side for the two vignettes were both above midline, while they clearly diverged significantly at the no-retaliation end. Choosing *not* to retaliate in the covert scenario (M=5.43, SD=1.92, 95%=[5.13 -5.73]) was significantly greater that failing to retaliate in the overt scenario, which participants clearly registered as overall blameworthy: (M=3.55, SD=1.99, 95%=[3.23-3.87]); t(302)=8.38, p<0.001, d=0.965). By contrast, on the retaliation side, the average rating of positivity in the overt scenario (M=6.70, SD=2.02, 95%CI=[6.37 -7.02]) was significantly greater than the average rating in the covert scenario: (M=5.65, SD=1.70, 95%CI=[5.37-5.93]); t(289)=4.87, p<0.001, d=0.685).

In exploring the interaction between agent and transparency, we were especially concerned with 1) how people rated retaliation for the friend in each vignette and 2) how people judged the actions of the person who failed to retaliate for himself—two of our most compelling and revealing results. Retaliating for one's friend in the overt, confrontation scenario (M=7.46, SD=1.69, 95%CI=[7.13-7.79]) was rated overwhelmingly higher than retaliating for one's friend in the covert, promotion scenario: (M=5.93, SD=2.12, 95%CI=[5.41-6.44]); t(92.8)=4.31, p<0.001, d=2.54). Meanwhile, *abstaining* from retaliating for oneself was rated much higher in the covert scenario (M=6.15, SD=1.58, 95%CI=[5.71-6.59]) than in the overt scenario: (M=4.12, SD=2.08, 95%CI=[3.72-4.52]); t(92.8)=5.06, p<0.001, d=1.09). It is also intriguing that choosing

not to retaliate for a friend in the overt scenario (M=3.32, SD=1.87, 95%CI=[2.95-3.69]) was rated lowest off all conditions in either scenario (along with not retaliating for a third party). Further, it was much different than the rating for the coinciding friend condition in the covert scenario: (M=5.49, SD=1.75, 95%CI=[4.97-6.00]; t(100)=6.04, p<0.001, d=1.19**)**. Together, these results seem to indicate that the more retaliation is endorsed, the higher people rate retaliation for a friend, and the more retaliation is censured, the higher people rate abstaining for oneself, in particular. However, this "covert" scenario seemed to suggest that people did not quite condemn retaliation. As stated, their response seemed more accurately to reflect ambivalence. Study 4 explores a scenario in which we expected people would regard retaliation as unequivocally negative.

Study 2b Discussion

Study 2b, particularly in conjunction with Study 2a, served to add considerable nuance to our hypotheses and further developed how social dynamics may affect perceptions of retaliating—and of course, the implications of choosing not to. The findings appeared to reflect a certain ambivalence—if not quite condemnation— of retaliation. People evidently considered the private nature of it a bit unsavory. The general pattern of results seems to support this interpretation. Retaliation findings were attenuated in comparison to Study 2a and there was no significant difference among the three agent conditions. Likewise, overall failure to retaliate was rated much higher than it was in Study 1 and Study 2, which we formally examined in comparing the "overt" and "covert" characteristics within the same model. Refraining from retaliation for oneself was rated the highest of the six conditions and this time, even significantly higher when compared to retaliation for oneself.

One likely explanation for these findings is that in the overt example, people may have regarded openly retaliating as reflecting more integrity, transmitting the idea that the Agent is moral and otherwise pro-social because he is willing to risk cost to himself to retaliate. For the same reason, one might expect choosing *not* to retaliate to be construed as an act of restraint or even magnanimity under covert circumstances because there is no audience to condemn retaliation.

As a final point, these studies suggest that retaliation also functions differently from mechanisms one might have thought comparable—say a straightforward scenario in which an Agent has the opportunity to choose to help. As mentioned, research has shown that—in between-subjects circumstances— acting to help a third party is rated as more moral and good than acting to help kin. In other words, retaliation seems to differ from other examples of assistance if it may be called so. In our next study, Study 3, we therefore, proposed to directly compare an instance of retaliation with one of more pure assistance.

STUDY 3: Retaliation vs. Help

Study 3 directly compared acts of retaliation and acts of help within the same varying social-relational context, this time using only friend and third-party comparisons. In accordance with previous research, we predicted that in the "help" context, providing assistance to third parties would be rated as more praiseworthy than those who helped a friend. As earlier discussed, studies have demonstrated that people not only afford special accommodations to friends (e.g., Marshall et al., 2020) and kin (e.g., McManus, Kleiman-Weiner, & Young 2020) but expect the same of others. In such studies, the authors found that participants rated people as more commendable when they helped strangers, theorizing that such behavior was considered beyond the call of duty to allies and family and, therefore, especially laudable.

However, in our previous studies, participants consistently rated retaliating on behalf of a friend—which might be conceived as a form of assistance— as more commendable than retaliating on behalf of a stranger, sometimes significantly but always in magnitude. Here, we, therefore, proposed to explore this discrepancy by directly comparing people's impressions of retaliating vs. helping a friend and stranger. Based on the body of literature outlined above and the addition of our results thus far, we predicted opposing patterns of results for the retaliation and help conditions, respectively: People would evaluate retaliating for a friend more positively than retaliating for a stranger and the reverse result in the case of the assistance scenario. This essentially constituted some test of our more speculative second hypothesis: that retaliation is essentially partial, even the particular mandate of allies.

Methods

Participants. 200 participants (*M*(age*)*= 39.3, *SD*=11.47) were recruited through Turk Prime for the study, completing the experiment in a median time of 3.38 minutes; 89 were female. They were compensated $0.40. None were excluded from analysvignette from Study 2 and an additional "assistance" scenario in which the Agent gives his after-hours office time to help a co-worker (friend or unknown, third party), who is struggling with a project. Participants read one of the following vignettes:

TABLE 1.13: Study 3—Wording for the "retaliation" and "help" conditions, respectively; the retaliation vignettes are almost identical to those used in Study 2a, except that the names and a few minor details have been changed. Agents acting as retaliators appear in red or as "helpers" in purple; friends appear in blue, third parties in green. Offenders appear in bold font.

Retaliation, Friend: Jordan and Alex are long-time co-workers at a company, who have become good friends over the years and mutually relied upon one another's help and support. They also work with **Casey**, though neither of them knows him very well. The three have recently been assigned to jointly lead a very lucrative project for an important client. The project involves a large team and much creative and arduous work. Technically the most senior of the three, Jordan has the authority to oversee the team and remove people at his discretion.	*Retaliation, Third Party*: Jordan, Alex, and **Casey** are long-time co-workers at a company; they don't know each other very well but have recently been assigned to jointly lead a very lucrative project for an important client. The project involves a large team and much creative and arduous work. Technically the most senior of the three, Jordan has the authority to oversee the team and remove people at his discretion.

After reading this portion, participants assigned to either of these retaliation conditions, read the following:

> At an early team meeting, **Casey** shocks Jordan by taking credit for a lot of Alex's work, mocking Alex each time he tries to state an opinion, and basically dictating the next steps he thinks Alex should be taking in the project.
>
> Jordan is mortified on Alex's behalf and knows **Casey's** behavior to be wrong. In front of the whole group, he sternly reprimands **Casey** and throws him off the project.

Helping conditions were worded as follows:

Help, Friend: JordaAlex are long-time co-workers at a company, who have become good friends over the years and mutually relied upon one another's support. Recently, each has been assigned to separate long-term projects involving important clients. They have both been under a lot of pressure and working long hours. One night, Alex is overwhelmed with work involving one particular subject for which he does not have a lot of familiarity; he has hit a serious snag and is struggling to get a document completed before an impending deadline.	*Help, Third Party*: Jordan and Alex are long-time co-workers at a company, who don't know each other well and have never exchanged more than a few words. Recently, each has been assigned to separate long-term projects involving important clients. They have both been under a lot of pressure and working long hours. One night, Alex is overwhelmed with work involving one particular subject for which he does not have a lot of familiarity; he has hit a serious snag and is struggling to get a document completed before an impending deadline.

At this juncture, the participants who read either of the “help” conditions, then read the following:

> Jordan happens to run into him as the two are breaking for coffee, and when Alex vents about his stress, Jordan decides to put aside his own work for the night and talk Alex through the major issues holding him up. While he has no stake whatsoever in Alex’s project, Jordan knows that he has far more expertise in such areas. With his help, Alex will be able to complete the document much more efficiently and effectively.

Participants then provided the same evaluations as in the previous studies: Counterbalanced measures of praiseworthiness/blameworthiness and goodness/badness of action, along with loyalty and friend ratings, all once again registered on bipolar 9-point scales. We also used the same behavioral measure that called on the participant to appraise how likely a particular agent would be to help in future. In a change from the previous study, however, we asked a different question about the likeliness of the agent engaging in future conflict: we instead posed the question, "People have various tendencies toward retaliation when they believe others have been unkind to them: Do you think that [Agent] is likely to retaliate against someone who has been unkind to him?" We included this measure in an effort to gauge what people might infer about help and retaliation when provided with information about one or the other.

Results

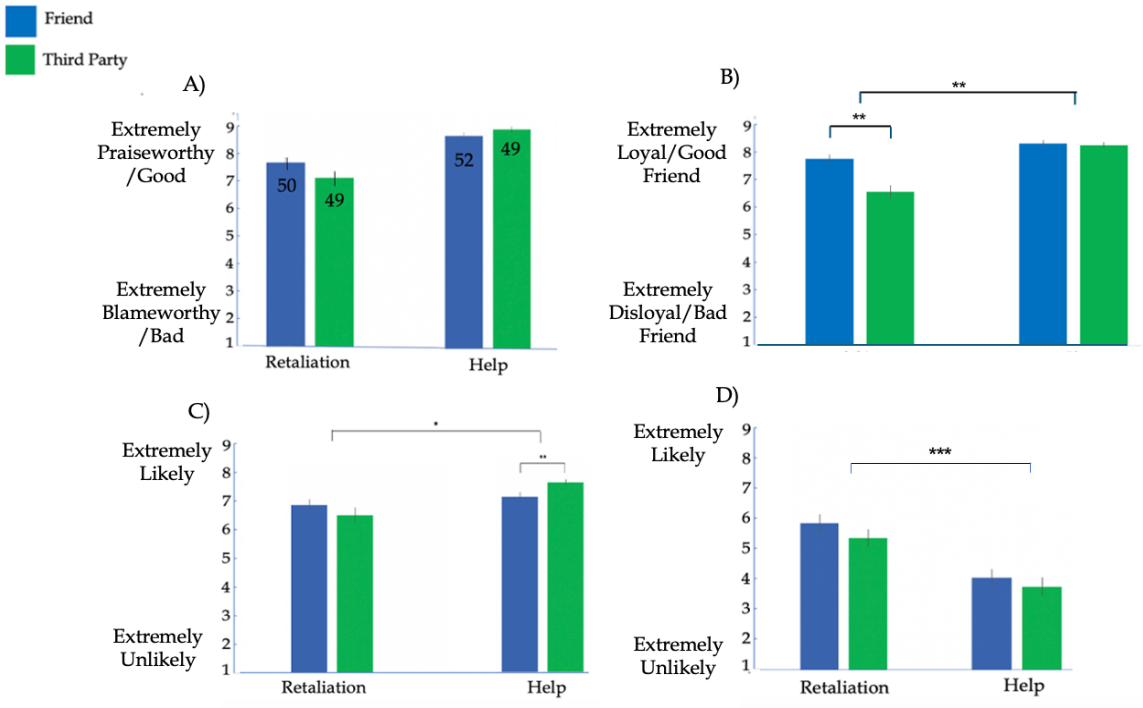

Figure 1.8. Study 3: A) Composite evaluations of praiseworthiness/goodness-of-action measures, B) Composite evaluations of loyalty and friendship measures, C) Answers to "People have various tendencies towards helping others in need: In general, do you think that Jordan is likely to help someone he observes to be in need?, and D) Answers to "People have various tendencies toward retaliation when they believe others have been unkind to them: In general, do you think that Jordan is likely to retaliate against someone who has been unkind to him?"—all as a function of agent identity (friend vs. third party) and action (retaliation vs. no retaliation). Bars represent *SE's.*

Praiseworthiness/Goodness of Action

Once again, finding the praiseworthiness and goodness of action measures to be highly correlated (r=0.850, p<0.001), we combined them into a single measure, which we will again refer to as "positivity."

As depicted in Table 12, the model produced a coefficient of determination, adjusted R^2=10.3, $F(3, 196)$= 10.3, p<0.001. Against our prediction, the main effect of agent was not significant, nor was the interaction between agent and action, which was just short of marginal: $F(1, 295)$=2.54, p=0.112, η_p^2 =0.013. There was a main effect of action: $F(1, 296)$= 23.4, p<0.001, η_p^2 =0.124.

TABLE 1.14--Study 3

Praiseworthiness/Goodness of action model for (retaliation vs. help) x agent(friend vs. third party)

POSTIVITY (Praiseworthiness & Goodness-of-Action)			
Effect/Measure	*F*	*p*	η_p^2
Coefficient of determination=R^2(adjusted)=0.123	$F(3, 196)= 10.3$	<0.001	----
Main effect of agent	$F(1, 296)=0.442$	0.507	0.002
Main effect of action	$F(1, 296)= 23.4$	<0.001	0.124
Full interaction: agent x action	$F(1, 296)=2.54$	0.112	0.013

The results trended in the predicted directions but only in magnitude. Retaliation for a friend (*M*=7.38, *SD*= 1.67, 95%CI=[6.88-7.81]) was not significantly greater than retaliation for a third party (*M*=6.82, *SD*=2.38, 95%CI=[6.34-7.30]); *t*(85.8)=1.26, *p*=0.210, *d*=0.255). Likewise, the third-party helper (*M*=8.42, *SD*=1.03, 95%CI=[8.13-8.71]) was greater in magnitude but not significance than the corresponding friend (*M*=8.21, *SD*= 1.41, 95%CI=[7.90-8.52]); *t*(98.8)=1.26, *p*=0.391, *d*=0.199). As might have been expected, adding loyalty to the positivity model produced a strong main effect: $F(1, 192)=40.0$, $p<0.001$, $\eta_p^2=0.626$, yet no interactions. Such a result suggests that people do not base their evaluations regarding helping behavior on perceptions of loyalty, in obvious contrast to our findings regarding retaliation.

Loyalty/Likeliness of Being a Good Friend—

Once again, finding the loyalty and likeliness of being a good friend to be sufficiently correlated (r=0.794, p<0.001), we combined the measures. The coefficient of determination for the model was R^2 (adjusted)=0.164, F(3, 196)= 14.0, p<0.001, as exhibited in Table 1.15 below.

TABLE 1.15: Loyalty/Friendship model

LOYALTY/FRIENDSHIP COMPOSITE	*F*	*p*	η_p^2
Coefficient of determination=R^2= (adjusted)=0.164	F(3, 196)=14.0	<0.001	---
Main effect of agent	F(3, 196)=8.71	<0.001	0.041
Main effect of action	F(3, 196)=26.3	<0.001	0.118
Interaction: agent x action	F(3, 196)=6.86	0.009	0.034

A two-way, factorial ANOVA yielded a significant interaction between the retaliation and help conditions: F(1, 196)=6.86, p=0.009, η_p^2=0.034, with the help conditions exceeding retaliation. There was an effect of retaliation, F(1, 196)=29.8, p=<0.001, η_p^2=0.13, in which the Agent acting on behalf of a friend (M=7.71, SD=1.33, 96%CI=[7.34-8.07]) was deemed more loyal and more likely to be a good friend than the retaliator acting on behalf of a stranger (M=6.52, SD=2.3096%, 95%CI=[5.88-7.16]; t(76.7)=3.14, d=0.635). There was no significant difference between the help conditions: M(friend)=8.25, SD=1.08, 95%CI=[7.96-8.54] and M(third party)=8.19, SD=1.03, 95%CI=[7.90-8.48]; t(99.0)=0.312, p=0.755.

Figure 1.9. Study 3—Pro-social measure:

"People have various tendencies towards helping others in need: In general, do you think that Jordan is likely to help someone he observes to be in need?

A two-way ANOVA provided a significant interaction between the retaliation and help conditions: $F(1, 196)=5.12$, $p=0.026$, $\eta_p^2=0.026$, with the help conditions rated more likely to "help someone in need" than the retaliation conditions, as further reflected in the main effect of action: $F(1, 196)=17.1$, $p<0.001$, $\eta_p^2=0.08$. There was also a significant effect between the help conditions, in which the Agent acting to aid a third party ($M=8.40$, $SD=1.33$, 95%CI=[8.18-8.62]) was judged more characteristically helpful than the Agent acting for a friend ($M=7.85$, $SD=2.30$; 95%CI=[5.88-7.16]; $t(84.7)=2.63$, $p=0.010$. While this is not an unexpected result, it is perhaps noteworthy considering the significance of the interaction between the help and retaliation patterns: $F(1,196)=5.17$, $p=0.024$. The retaliation conditions did not differ but trended in the other direction, with the actions of the retaliator acting for a friend rated at ($M=7.56$, $SD=1.49$, 95%CI=[7.15-7.97]) and those of the third-party retaliator rated at $M=7.16$, $SD= 0.300$, 95%CI=[6.57-7.74]); $t(86)=1.08$), $p=0.281$.

Figure 1.10. Study 3—Deterrence measure:

"People have various tendencies toward retaliation when they believe others have been unkind to them: Do you think that [Agent] is likely to retaliate against someone who has been unkind to him?"

A two-way ANOVA failed to produce an interaction: $F(1, 196)=0.1225$, $p=0.727$, $\eta_p^2<0.001$), yet there was a main effect of action: $F(1, 196)=20.1$, $p<0.001$, $\eta_p^2=0.093$), with retaliation judged to be more of a deterrence than helping behavior. There was no main effect of agent—which is noteworthy, given the interaction for pro-social behavior.

In both cases—retaliation and help—the friend condition was judged to be more inclined for retaliation, though neither comparison was significant.

Study 3 Discussion

Although the agent by action interaction for praiseworthiness was not significant, it trended in the predicted direction, which is, in fact, consistent with some of the small effect sizes that helping behavior (when between-subjects) has produced in other studies (e.g., McManus et al., 2020). Furthermore, in conjunction with the results from the other DVs, this experiment tended to lend support to our position that retaliation is not a sub-category of assistance and operates differently in predictable ways. Retaliation for another—particularly an ally— seems to signal that one is likely to defend one's own interests against exploitation of this kind. The interaction in the pro-social measure suggests that retaliating for an ally compared to a third party may serve to demonstrate a more expansively helpful nature—e.g., not limited to responding to an offense. As expected, the helping conditions exhibited the opposite relationship with third parties deemed more characteristically helpful than friends. Further, the helping behavior seemed to stay squarely within that realm. In other words, it predicted future help, but did not seem to suggest anything at all about their likelihood of retaliating, for instance. There was no difference between the two conditions, and they were rated very low in overall likelihood of future retaliation. Likewise, loyalty did not seem to factor into the helping behavior—or distinguish between friend and third party—as it has so reliably with retaliation. Collectively these patterns seem to contribute to our general findings that retaliation may be the particular purview of allies and that it may represent a qualitatively different heuristic or mechanism than both assistance and third-party retaliation.

One of our most robust and compelling findings is that retaliation for an ally is consistently rated more positively than for retaliating for oneself but also for a third-party stranger. While retaliation for a friend has not always been *significantly* greater than retaliation for a third party, it has always had a greater magnitude (mean). Notably, in the covert scenario, in which people seemed to register greater ambivalence about that retaliatory act, itself, the ally condition retained its greater magnitude but lost its significance. In a follow-up study further exploring retaliating and helping behavior, we therefore, considered how a truly hostile, incongruent, and disproportionate act of retaliation would affect these social dynamics.

STUDY 4*: Retaliation vs. Helping—Gossip*

Study 4 had two primary objectives: First, to explore another case of "helping" behavior and to determine if the results would tend to align with our findings from the previous study. Second, we created an office scenario that we expected participants to condemn. If participants did not approve of the mode of retaliation, we predicted that they might no longer favor retaliation for allies and might even rate third -party retaliation higher, as the nature of the act might undermine any obligation to engage in it. Further, the third party would have, in this case, the distinct advantage of appearing less self-interested. We speculated that the reprehensible nature of the retaliation might convert appraisals of high-minded loyalty to mean-spirited favoritism. However, as Lieberman & Linke (2007) found, people applied less punishment for offenses to kin and allies while still condemning their action to equal degree**.**

Methods

Participants. 204 participants (*M*(age)= 40.8, *SD*=10.5) were recruited through Turk Prime for the study, completing the experiment in a median time of 6.16 minutes; 80 were female. They were compensated $0.40. None were excluded from analyses.

Design & procedure. As noted, we employed a 2(Agent: friend vs. third party) x 2(Action: Retaliation vs. Helping) between-subjects design, using what we expected would be considered an egregious act of retaliation and an additional "assistance" scenario in which the agent devotes his after-hours office time to help either a friend or an unknown co-worker with a project he is struggling with. Participants read one of the following vignettes:

TABLE 1.16 : Study 4 retaliation and help vignettes: Agents acting as retaliators appear in red or as "helpers" in purple; friends appear in blue, third parties in green; offender printed in bold font

Retaliation, Friend: Taylor and Jordan are long-time employees in the marketing department of Company X who have become good friends over the years. **Derek** also works in the department, though neither knows him well and have never really exchanged more than a few words in passing. Recently, all three have been assigned to a large team for a potentially lucrative project attached to an important client. In the course of completing the project, the team has had to collaborate and discuss details over many meetings. Before long, it becomes clear that **Derek** has a tendency to single Jordan out for bad feedback and ridicule; in meetings, he constantly interrupts Jordan, dismisses his ideas, and ends up sticking him with the most menial of tasks.	*Retaliate, Third Party:* Taylor, Jordan, and **Derek** are long-time employees in the marketing department of Company X. They don't know each other very well and have never really exchanged more than a few words in passing. Recently, all three have been assigned to a large team for a potentially lucrative project attached to an important client. In the course of completing the project, the team has had to collaborate and discuss details over many meetings. Before long, it becomes clear that **Derek** has a tendency to single Jordan out for bad feedback and ridicule; in meetings, he constantly interrupts Jordan, dismisses his ideas, and ends up sticking him with the most menial of tasks.

From here, each participant assigned to one of the two retaliation conditions read the following:

TABLE 1.16 (continued):

By the end of a few weeks, Taylor has become exasperated and angry with what he has witnessed of **Derek's** behavior towards Jordan. One day, Taylor is busy compiling a report from documents that he has received from various members of the team when he comes across a private email chain **Derek** has passed along accidentally. The chain contains very personal emails between **Derek** and his wife, suggesting that **Derek** is having an extramarital affair.

That night, Taylor meets up with some co-workers for cocktails. When **Derek**'s name comes up, he tells the group that he has it on good authority that **Derek** has been cheating on his wife.

Participants in the "help" conditions read one of following:

Help, Friend: Taylor and Jordan are long-time employees in the marketing department of Company X, who have become very good friends over the years. Recently, Jordan has been assigned to lead a potentially lucrative project for an important client. At the start of the project, Jordan orders all of the background files to be sent over on a Saturday and sets aside that day to receive them and start going through them.	*Help, Third Party:* Taylor and Jordan are long-time employees in the marketing department of Company X. They don't know each other very well and have never really exchanged more than a few words in passing. Recently, Jordan has been assigned to lead a potentially lucrative project for an important client. At the start of the project, Jordan orders all of the background files to be sent over on a Saturday and sets aside that day to receive them and start going through them.

At this point, all participants assigned to one of the two "helping" conditions each read the following.

That Saturday, the files are delivered in the lobby, and he discovers that the load is much bigger and in greater disarray than he has supposed. None of his team and few admin are around to assist as the building is mostly empty for the weekend.

Just as he is starting to load the first batch onto the elevator, he runs into Taylor, who had stopped by work to catch up on some of his own projects and is now headed home. Taylor sees what Jordan has on his hands and offers to help get the boxes upstairs. Though it is not at all his job to assist in this way, and he has nothing to do with this project, Taylor spends a 1/2 hour or so helping Jordan to get the materials in manageable order.

After randomization into one of the four conditions, participants completed exactly the same procedure as in the previous study. They first answered counterbalanced questions regarding praiseworthiness, goodness of action, propensity to be loyal, and likelihood of being a

good friend. Once again, the four measures were found to be highly correlated and combined into two. Participants then completed the same two counterbalanced behavioral questions from Study 3, one targeted at inferences regarding pro-sociality and the other at deterrence—or exhibition of strong character, unlikely to tolerate negative behavior towards themselves.

Results

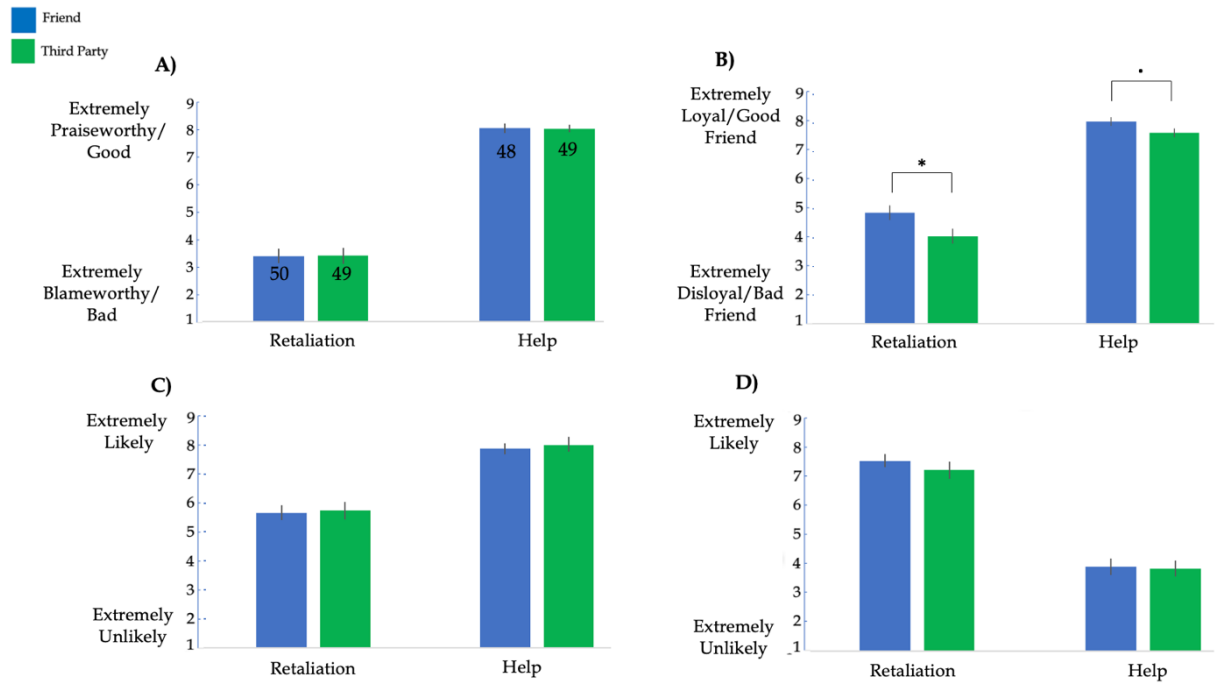

Figure 1.11. Study 4: A) Composite evaluations of praiseworthiness/goodness-of-action measures, B) Composite evaluations of loyalty and friendship measures, C) Answers to "People have various tendencies towards helping others in need: In general, do you think that Agent is likely to help someone he observes to be in need?, and D) Answers to "People have various tendencies toward retaliation when they believe others have been unkind to them: In general, do you think that Agent is likely to retaliate against someone who has been unkind to him?"—all as a function of agent identity (friend vs. third party) and action (retaliation vs. no retaliation). Error bars represent *SE's*.

Praiseworthiness/Goodness of Action

Once again, finding the praiseworthiness and goodness of action measures to be highly correlated (r=0.926, p<0.001), we combined them into a single measure of positivity. As exhibited in Table 17, adjusted R^2=0.690, $F(3, 186)$=142, p<0.001, the model was significant, but

because of the nature of the results, there was virtually nothing to report about the main effects that was enlightening, except that all the variance is concentrated in action, which was significant at $F(1, 286)= 209$, though with a miniscule effect size. There was no interaction.

TABLE 1.17: Study 4

Positivity model for action(Retaliation vs. Help) x agent(Friend vs. Third Party)

POSTIVITY (Praiseworthiness + Goodness-of-Action) *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2 (adjusted)=0.690	$F(3, 186)= 142$	<0.001	----
Main effect of agent	$F(1, 186)=0.00$	0.983	<0.001
Main effect of action	$F(1, 286)= 209$	<0.001	<0.001
Full interaction: agent x action	$F(1, 186)=0.002$	0.996	<0.001

Positivity for the two conditions of retaliation and helping behavior, respectively, were similar. As expected, people rated retaliation very negatively for the first time, with retaliation for friend at M=3.40, SD=1.82, 95%CI=[5.83-6.97] and retaliation for third party at M=3.41, SD=2.00, 95%CI=[2.85-3.97]. Unfortunately, the vignette we employed for assistance may once again have encountered ceiling effects, as helping a third party (M=8.10, SD=0.918, 95%CI=[7.83-8.36]) was no higher than aiding a friend, (M=8.06, SD=1.17, 95%CI=[7.73-8.39]).

Loyalty/Being a Good Friend

Yielding a correlation of r=0.814, p<0.001, we once again combined general likeliness of being loyal and a being a good friend into one combined measure. As exhibited in Table 18, R^2= 0.577 was significant for the model at $F(3, 186)=84.6$, p<0.001.

TABLE 1.18—Study 4

Loyalty/Friendship model: action (retaliation vs. help) x agent(friend vs. third Party)

A) LOYALTY/FRIENDSHIP COMPOSITE *Effect/Measure*	*F*	*p*	η_p^2
Coefficient of determination=R^2(adjusted)=0.577	$F(3, 186)= 84.6$	<0.001	----
Main effect of agent	$F(1, 186)=1.53$	0.217	0.008
Main effect of action	$F(1, 286)= 136$	<0.001	0.423
Full interaction: agent x action	$F(1, 186)=1.02$	0.315	0.005

There was no significant interaction. Despite its low positivity rating for both conditions, retaliation nonetheless produced another intriguing result: Though judged equally blameworthy, the retaliator acting for a friend was rated at, (M=4.88, SD=1.80, 95%CI=[4.38-5.38]), which was significantly higher than its third-party counterpart: M=4.05, SD=1.89, 95%CI=[3.52-4.58]; $t(94.8)=2.21$, p=0.030, d=0.449. Likewise, the friend who helped (M=8.05, SD=1.03, 95%CI=[7.76-8.34]) was rated as marginally more loyal than the third-party helper (M=7.67, SD=1.03, 95%CI=[7.36 -7.98]=; $t(90.3)=1.79$, p=0.077, d=0.371).

Figure 1.12—Pro-social measure:

"People have various tendencies towards helping others in need: In general, do you think that [Agent] is likely to help someone in need?

As expected, people inferred that those who engaged in helping behavior were judged to be more likely to do so in future, and there was no difference between friend (M=7.88, SD=1.30, 95%C1=[7.51-8.25]) or third-party conditions (M=8.00, SD=1.51, 95%C1=[7.56-8.44]). More notably, despite the extremely blameworthy nature of both retaliator's actions, people seemed to infer that they would hypothetically assist others or at least they were both narrowly rated above the midpoint, friend (M=5.67, SD=1.74, 95%C1=[5.18-6.16]; $t(47)$=4.64, p<0.001) and third party (M=5.73, SD=2.0495%C1=[5.22-6.24]; $t(48)$=4.24, p<0.001). As will be further demonstrated below, people apparently presume helping behavior from extreme retaliatory behavior, though not the inverse.

Figure 1.13. Study 4—Deterrence measure:

"People have various tendencies toward retaliation when they believe others have been unkind to them: Do you think [Agent] is likely to retaliate against someone who has been unkind to him?

As with the other DVs, there was a clear effect of action ($F(1, 186)$=136, p<0.001, η_p^2=0.433) but no interaction between agent and action. In this case, this effect is perhaps worth noting because we were especially interested in how people would respond to retaliation for another—friend or third party— when it was regarded as repugnant or disproportionate. Again, despite clearly finding the manner of retaliation to be blameworthy and arguably unwarranted, people rated both retaliators highly on this measure, indicating that retaliation did transmit to onlookers a proclivity for retaliating for *oneself*—a kind of deterrence signal. There was no

significant difference in this measure between retaliating for a friend, (M=7.5, SD=1.56, 95%CI=[7.06-7.94]), and for retaliating for a third party (M=7.18, SD=2.00, 95%CI=[6.74-7.75]); $t(90.3)=0.868$; $p=0.388$, $\eta_p^2=0.102$. In addition, comparing Study 3's and Study 4's retaliation results (each group of which was between-subjects) for this measure, there was a clear effect of study: $F(1, 192)=22.0$, $p<0.001$ with people anticipating that those who engaged in the objectionable scenario (Study 4) much more likely to engage in retaliation for themselves than their Study-3 counterparts. Pair-wise t-tests revealed that the agents who retaliated for a friend or third party in Study 4 were considered much more prone to retaliation for themselves than the same two conditions in Study 4. The friend in study 3 was rated significantly less likely to retaliate than the friend in Study 4: ($t(89.4)=4.50$, $p<0.001$). Likewise, the difference in third parties produced: ($t(96.0)=4.51$, $p<0.001$). Once again, as in Study 4, people did not appear to infer retaiatory characteristics from helping behavior, nor did they differ from one another. Helping a friend produced a likelihood of retaliating at M=3.88, SD=1.93, 95%CI=[3.33-4.43] and helping a third party: M=3.82, SD=1.83, 95%CI=[3.27-4.37].

Study 4 Discussion

The data provided evidence for our primary prediction for Study 4: that people would rate the retaliation scenario negatively and, we hypothesized, would no longer favor retaliation for friends for this reason. Interestingly, loyalty was still rated higher for the friend condition. In addition, people both inferred helping behavior from even this negative retaliatory action and also the tendency to retaliate on their own behalf, which utterly eclipsed the same results from Study 3. As one final point, this mode of retaliation—though severely condemned—nonetheless evidently impressed people with the idea that these Agents were not to be trifled with. However,

such gains may be rather short-term, and may run the risk of alienating people who might regard the retaliation as excessive or frightening.

TABLE 19: Brief overview of studies

Study 1: Park Theft

--Retaliation overall rated highly

--On the retaliation side, the friend condition was greatest in magnitude, not significance

--On the non-retaliation side, failure to retaliation for a friend was most blameworthy

--PCA analyses revealed that, on the retaliation side, the friend was most "pro-social" than wither of the other two conditions; retaliation for oneself was lowest.

--On the retaliation side, the inverse was true

--Loyalty was both a significant predictor of praiseworthiness and interacted with condition to exert an additional influence.

Study 2a: Overt scenario: Office confrontation

--People found retaliation to be much more praiseworthy than non-retaliation; notably, abstaining from retaliating for oneself was also rated highly and above the midline

--On the retaliation side, the friend condition was dominant, significantly greater than both the self and third-party conditions

--While it, once again, trended in the predicted direction, the third-party retaliation condition was greater than the retaliatory self condition only in magnitude

--On the non-retaliation side, the self condition was much higher than both friend and third-party conditions, which were decidedly blameworthy and about even in magnitude

TABLE 1.19 (continued):

--Once again, loyalty was both a significant predictor of praiseworthiness and interacted with condition, eliminating the effect of action and diminishing the effect of agent

Study 2b: Covert scenario: Office promotion

--People seemed to regard retaliation with ambivalence, both retaliation and non-retaliation deemed moderately praiseworthy: collapsed across action, there was no difference between retaliation and no retaliation

--On the retaliation side, the friend was greater only in magnitude; there were no significant differences among the three.

--On the non-retaliation side, once again abstaining from retaliation for oneself was higher than the other two conditions; notably, in this case, refraining from retaliation was *higher* than retaliating for oneself

--Loyalty, once again, was a significant predictor and interacted with condition when regressed on praiseworthiness

--Direct comparison of the two models (overt vs. covert) further supports:

--Retaliation and non-retaliation are not simply the inverse of one another

--When retaliation is overall endorsed, retaliation for allies is dominant, suggesting it may be the domain of allies

--When people rate retaliation ambivalently or more negatively, abstaining from retaliation becomes much more praiseworthy

Study 3: Covert, office confrontation retaliation vs. Help with difficult work

--Regarding praiseworthiness, while the interaction was not significant, it was near marginal and trended in the expected direction, with retaliation friend rated higher than for third party; as expected, the opposite was true of the "help" conditions.

TABLE 1.19 (continued):

--Once again, loyalty predicted praiseworthiness, with loyalty for a friend higher than loyalty for a third party. There was a significant interaction with the "help conditions," for which loyalty did not differ.

--Helping measure : On the helping side, the third-party condition was rated significantly more likely to aid someone in need than the friend condition. There was also an interaction. Both retaliation conditions were also rated highly, with friend higher than third party in magnitude. Interestingly, the four conditions did not differ much. People clearly inferred future helping behavior from present retaliatory behavior.

--Retaliation measure: The retaliation conditions were higher—though they weren't extremely high (in comparison to Study 4's findings). The helping conditions were well below the midline

--Together, these findings demonstrate that retaliation can signal future helping behavior, but the reverse does not seem to be true

Study 4: Vindictive revenge vs. Help with moving boxes

--Retaliation rated as very blameworthy; the help conditions were rated extremely highly and were almost dead even, perhaps due to ceiling effects

--Loyalty: the helping conditions were rated much more highly; for both retaliation and helping behavior, the friend was significantly rated as generally more loyal than the third party

--Helping-behavior measure: the two helping conditions did not differ but were rated very highly. Notably, people also apparently inferred likelihood of helping others from severe retaliation because the retaliation conditions were also above the midline.

TABLE 1.19 (continued):

--Retaliatory tendencies: Despite being rated as blameworthy, the retaliatory conditions clearly signaled a harsh, high likelihood of tending to retaliate for oneself (much higher than the same conditions in Study 3). This was not true of the helping conditions, who were rated as very unlikely to engage in retaliation for oneself.

--Both Studies 3 and 4 demonstrate that retaliation and help operate differently and signal different attributes

GENERAL DISCUSSION—

The study findings provide strong evidence that people factor in both the identity of the retaliator and, more specifically, the relationship between retaliator and victim when evaluating decisions to retaliate—or not. Collectively, our studies encompass a few overarching themes and findings: First, evaluations of retaliation (or choosing not to)—whether for self or others— can be positive or negative or anywhere in between. This evidence suggests it is volatile and intensely subject to other circumstances, such as open confrontation or transparency. Second, as predicted by both of our first two Hypotheses 1a and 1b—Pro-sociality and Integrity, respectively—retaliating for others was generally regarded more positively than for oneself in magnitude but only sometimes in significance. In contrast, deciding *not* to retaliate for oneself appears to be a positive reputational signal—especially when overall retaliation is not considered particularly laudatory. This result, in particular, also suggests that retaliation is a valuable resource, to be used judiciously. Third, People value transparency and proportionality in retaliation, demonstrating ambivalence and differing patterns of praiseworthiness towards those who retaliate privately or incongruently with the original offense. Fourth, the results suggest that

retaliation is both inherently partial in addition to being a valuable, costly resource. For one, loyalty featured prominently in praiseworthiness evaluations; for another, people tended to regard retaliating for oneself as less favorable than retaliating for others, suggesting that it is regarded as comparatively selfish and, therefore, desirable. By this logic, overall retaliation is valuable. Fifth, retaliation does *not* appear to be a form of assistance, as it displays different social-relational patterns from "helping behavior" in other studies and in our own Studies 4 and 5. Sixth, retaliation for allies seems to manifest a particular prerogative or mandate, being consistently higher in praiseworthiness than other forms of retaliation but particularly when retaliation has been sanctioned. Once again, this does not preclude the possibility that retaliation for third parties is an independent mechanism, valuing impartiality.

To our knowledge, this is the first paper to directly compare evaluations of retaliation vs. non-retaliation, in addition to comparing categories of retaliators—especially as concerns valence. While a rising body of developing research—especially involving economic games and third parties—has examined participant appraisals of retaliators, they have not often directly compared retaliator identities or overall approval of the retaliating act. For instance, various experiments have suggested that third parties are frequently deemed impartial and trustworthy. However, the current research suggests that such characteristics may not always translate into praiseworthiness or endorsement for of retaliation—impartiality in particular. As discussed, we, in fact, argue that retaliation may be an inherently *partial* mechanism, at least as it applies to allies. After all, retaliation on behalf of friends was rated highest among the conditions in magnitude and non-retaliation of friends was considered extremely blameworthy. Yet, we do not exclude the possibility that third-party retaliation may be a separate mechanism altogether, having its own parameters.

Perhaps the simplest of our findings is that overall retaliation can be deemed positive, as best exhibited by Study 1 and Study 2a. In both those studies, all three retaliators (self, friend, and third party) were rated favorably and failing to retaliate for them rated as very blameworthy. The possible exception, of course, was choosing not to retaliate for oneself, which was rated at about the midline in Study and well into praiseworthiness in Study 2b. Furthermore, condoning retaliation through high praiseworthiness ratings produced a distinct pattern of results among the decision-makers—retaliators and non-retaliators. This pattern seemed to be predictably different when retaliation was regarded more ambivalently and, in turn, negatively. As represented most clearly by Study 2a (overt scenario)—in conjunction with Study 2b (covert scenario)—the greater the overall rating of retaliation, the more the friend condition emerged as most praiseworthy. Failing to retaliate for a friend was also lowest, but this time (unlike in Study 1) failure to retaliate for a third party was rated nearly as blameworthy.

Conversely, Study 2b, in particular, suggested that the *less* retaliation was endorsed, different patterns asserted themselves. Retaliating for a friend was not so dominant—greater only in magnitude, not significance. Critically, as retaliation became less laudatory and more ambivalent, abstaining from retaliating for oneself was rated higher. This finding seemed to support the Integrity Hypothesis, which asserts that choosing *not* to retaliate for oneself might represent a demonstration of magnanimity and/or transmit a message that this person has no need for retaliation. For instance, imagine a simple example in which someone has overheard a co-worker making her the butt of a joke. Displaying anger and retaliating in some way might make her appear petty and small-minded. Laughing it off might convey to others that she is secure in himself and not affected by such behavior or this mild "offender." Indeed, FeldmanHall and colleagues (2014) found that victims were less punitive than third parties in an economic games

that allowed for many options, including rectifying unfair divisions of money or rectifying *and* retaliating for unfair offers. There are, of course, circumstances that call for retaliation for oneself—as exhibited by the results of Studies 1 and 2a, in which it was rated highly. However, the material point is that choosing not to retaliate for oneself presents almost the unique opportunity—among all the conditions— to signal that one is "above" retaliation. In addition, the fact that choosing not to retaliate for oneself can apparently be selfless further reinforces that retaliation is costly—sometimes dangerous. It's therefore a valuable resource. To refrain from using it for oneself and reserve it for others is thus praiseworthy, all else being equal.

With regard to the Pro-sociality Hypothesis, there was some evidence that people tend to favor retaliation for others over self but not very consistently. While retaliating for oneself never exceeded retaliating for a third party, they were only significantly different in the PCA analysis when all measures were combined into a single DV. However, when considered in combination with the non-retaliation results, there were interactions. These interactions indicate that overall retaliating for oneself was perceived to be more blameworthy and probably because more selfish, as discussed. The most reliable finding from this standpoint—and in general— was that retaliation for a friend was most praiseworthy, primarily in magnitude. In fact, it was only in Study 2a and the PCA analysis in Study 1 that retaliation for a friend was *significantly* more praiseworthy than its two counterparts. As repeatedly discussed, however, we expect that retaliation for allies is much more likely to be rated highly when overall praiseworthiness is condoned, as it overwhelmingly was in Study 2a.

This result is consistent with our theory that retaliation must be considered *warranted* to activate this apparent heuristic that supports retaliation for allies. More specifically, the studies suggest that two features, transparency and proportionality between offense and retaliation, are

central considerations in whether retaliation is deemed justifiable. Study 2a was arguably the most noble and proportionate case of pure retaliation, and retaliation for allies was clearly predominant. Importantly, it was "warranted" by both standards, transparency and proportionality. The overt scenario (Study 2a) was open and visible to the offender and relevant bystanders, as well as being consistent with the original transgression: The Agent confronted the offender about his poor behavior in front of the group members and retaliated by taking him off a project about which he had been difficult and inappropriate. In contrast, we predicted and found that the scenario in Study 2b that people rated overall retaliation lower and failure to retaliate higher, indicating a kind of ambivalence. We expected that people may have construed a given retaliator's actions as surreptitious or sly and reflecting a reluctance to involve himself. The way the scenario was worded also suggested that the retaliator may never reported the incident or taken any action if he hadn't been coaxed by heavy questioning from the boss. Further, the mode of retaliation did not appear to be congruent with the original offense. As further evidence, in Study 4, retaliation was designed to be both harsh and incompatible with the instigating "offense." As predicted, people judged this retaliation to be very blameworthy and retaliating for a friend was, for the first time, not rated highest in magnitude.

Comprehensively, these studies thus strongly indicate that social-relational dynamics are highly relevant in retaliation. Perhaps most importantly, the collective findings seem to support our hypothesis that retaliation is the domain and mandate of allies. Other scholars have suggested such a possibility (e.g., McCullough et al., 2013), though few have directly tested it, as we have. As one of our most speculative hypotheses, it seems to provide the most consistent evidence. Friends were always rated highest in magnitude (with the exception of Study 4), and perhaps

most importantly, was increasingly robust the greater people praised overall retaliation. Failure to retaliate for friends was also generally rated lowest (with the exception of Study 2b).

The loyalty results—including the interactions in the first three studies—also argue once again that people prioritize retaliation for friends, and that when most praiseworthy, it is likewise most partial. Furthermore, in the PCA analysis in Study 1, retaliation for a friend was rated significantly higher than the other two; likewise, on the non-retaliation side, it was rated significantly lower than its two counterparts. Retaliation by third parties clearly occurs, but it seems to represent a different mechanism, actuated by different inputs. This possibility does not interfere with our position that retaliation may be, in a sense, the special purview of kin and allies, as some have asserted and our studies support. Favoritism for friends is not condoned in many contexts (e.g., Shaw et al., 2018), yet retaliation may be one psychological realm in which a kind of partiality is not only tolerated but, to a degree, lauded and expected to the exclusion of outsiders.

In support of this possibility, research has demonstrated that people regard the avoidance of inflicting harm differently than they do acts of assistance. As briefly discussed, harm inflicted in interpersonal relationships is much more indicative of trait morality and character than harm inflicted between strangers; for instance, a man who harms his wife or friend is regarded as more reprehensible—and *characteristically* reprehensible—than a man who harms a stranger (e.g., Hughes et al., 2016). In this way, assistance and harm are not consistent. People may regard retaliation to be in the "harm" domain; in a sense, responding to an outrage against a friend may be like a belated deflection of harm—or an act to prevent harm in future—and not merely an act of assistance. People may believe that it is more reflective of character to prevent harm on behalf

of friends than on behalf of strangers. However, as repeatedly noted, the retaliatory act, itself, appears to require approval for these dynamics to be in play.

Further, our studies comparing retaliation and help suggest that they function differently, as earlier discussed. Most obviously, retaliating for a friend is generally more praiseworthy than for a third party and the opposite is true of assistance. Nor are the two simple inversions of one another. For instance, we found that retaliation predicted future "helping behavior" while the opposite was not true. Relatedly, other studies have demonstrated that people not only value kin and even friends above strangers—including other ingroup members—but there is also evidence that they *expect* others to treat ingroup members more favorably and that they consider inflicting harm on a group member as especially noxious (Hughes et al., 2016). This is, of course, contrary to the research that, as discussed, has demonstrated that when it comes to many forms of assistance, people consider it to be more commendable across many dimensions of morality to help strangers than to help allies (e.g., McManus et al., 2019**).** Harm, may, therefore, be a central component in people's appraisals of retaliation in various forms. As noted, research suggests that people make particularly strong trait and action inferences—especially in the moral domain—when harm is involved. To use another example, another study showed that even lying to a spouse or friend was considered far worse than lying to a stranger (DePaulo & Kashy, 1998). In short, this rationale may indicate that failing to retaliate on behalf of a friend may be construed as a form of harm against that friend.

Similarly, retaliation almost inherently involves the infliction of harm on someone else to achieve what may be perceived as even an ultimate good: the future protection of the victim against exploitation. People may believe that such an act of aggression requires a certain prerogative or justification—in this case, the justification of the closeness of the relationship;

people may believe the closeness of the relationship—the friendship—better mitigates what would be a harmful, though warranted act: enacting aggression or violating a norm even for a good cause (e.g., Fiske & Rai 2014). However, these dynamics might change if one were retaliating against a close friend or associate. This factor was not manipulated here but would be of interest in future studies.

If these explanations have traction, the justifiability of the actions must be relevant. Future areas of investigation might include manipulating whether the various retaliators are in positions of authority and, in which case, a third-party might be vested with a different power—the power of being something of an *institutional* third-party rather than just a bystander (e.g., Marshall et. al 2020). Evidence already suggests that people tend to believe that third parties (e.g., Chen & Ma, 2020) are more impartial so perhaps with the proper power or authority, "retaliation" would seem more their jurisdiction and, therefore, more praiseworthy.

Other future directions might include exploring how people would evaluate friends committing some kind of moral or norm transgression in order to aid a friend or stranger. As discussed, Lieberman & Linke (2007) found that that people applied less stringent punishment to allies than to strangers for the same transgression, even when the act itself was rated equally immoral. Extending such findings, one might imagine constructing a study in which a person lies to protect a friend vs. a stranger who has stolen, particularly in a moment of desperation. Another provocative possibility might be to explore people's willingness to commit various transgression—a norm violation or even a criminal act—themselves, on behalf of a friend or ally versus a stranger and under what circumstances.

Indeed, much evidence has indicated that third parties—especially on an institutional level— have undoubtedly contributed to the maintaining of order in small ancient groups and

modern times (e.g., Burt & Knez, 1995) It seems, however, that retaliation for allies may constitute a separate mechanism from third-party retaliation or "punishment" altogether—especially given that impartial third parties or institutions are not and have not always been available, able, or willing to interfere (e.g., McCullough et al., 2013). Our results, perhaps, provide some insight into this possible separate nature.

Overarchingly, we set out to explore how identities and relationships affect how people evaluate retaliation. Such a pursuit is significant because it is precisely onlookers—in addition to offenders—who stand to judge retaliation and those who engage in it by conferring benefits or administering punishment. In investigating retaliation—or the decision not to—for self, friend, and third party, we found that it can be volatile and distinct patterns emerge the more people endorse or censure it. As we have discussed in detail, our findings suggest that retaliation seems to be a partial, valuable mechanism—perhaps the distinct purview of allies—and is a separate from both third-party retaliation and straightforward assistance.

After considering how people consciously evaluate retaliation, we turn to an example in which automaticity is likely much more in operation: the assessment of incidents involving slurs. In many ways the precise power of slurs remains murky, with perceived threat of the minority group often invoked (e.g., Scaptura, 2019). There's no question that the psychology surrounding the use and power of slurs is complicated. We set out to investigate how people would respond to the use of slurs—against white women, black men, and black women, respectively— and what kind of inferences these simple incidents would invoke in onlookers (participants). We centered the studies surrounding the use of misogyny, which lies at the intersection of some of the world's worse bigotry but closest relationships.

CHAPTER 2:
BIGOTRY LIGHT—EVIDENCE THAT MISOGYNY IS REGARDED AS LESS THREATENING AND SERIOUS THAN BLACK RACISM, AND IN FACT, HAS A DILUTING EFFECT ON ITS PERCEIVED NEGATIVITY

INTRODUCTION

When habitus encounters a social world of which it is the product, it is like a "fish in water": it does not feel the weight of the water, and it takes the world about itself for granted…"

--Pierre Bourdieu, *An Invitation to Reflexive Sociology*

Misogyny is ancient, originating, presumably, from the moment that a man asserted his superior physical strength over a woman and extended this disparity to power dynamics that have infected almost every social system around the world. We write "infected" because it is an infection of sorts, as are all forms of chauvinism in that they commandeer adaptive systems—e.g., heuristics— and pervert them into detrimental, erroneous thoughts and behavior—e.g., stereotypes and attitudes about groups that overwhelm individuation or assessments of people as *individuals* (Greenwald & Banaji, 1995). There is very little evidence that intellectual gender differences exist and what has been reported is highly suspect (e.g., Rudman & Glick, 1999), and likely a product of socialization and known psychological phenomena like "spheres of influence" (e.g., Almy & Sanatullova, 2016), stereotype threat, (e.g., Gunther et al., 2010) or what has been termed belonging uncertainty (e.g., Spencer et al., 1999; Smith; Shih et al. 1999). Yet, the extent of misogyny is nonetheless staggering, permeating all aspects of women's—and also men's—lives (e.g., Momennejad et al., 2019). As the oldest form of bigotry, it is also arguably the most perniciously stubborn and sluggishly resistant to reform, frequently battling for basic recognition: it is the water around Bourdieu's "fish," so powerful and insidious that we don't even recognize the nature of the medium that surrounds us.

The current research thus proposes to explore how people perceive and evaluate misogyny in comparison to another form of bigotry, anti-black racism. Our overarching hypothesis is that people regard misogyny as much more acceptable than racism and less threatening to the individual, the group, and society at large. Broadly, we contend that this

discrepancy is wildly untrue: that misogyny is very dangerous precisely because it is so deeply woven into the social fabric. It wreaks underestimated havoc on people for reasons that may ultimately reside largely in spurious and unsubstantiated beliefs with regard to biology and evolution—and a failure to account for differentiated socialization that begins in infancy (e.g., Campbell et al., 2000). Men and women, for one, tend to look categorically different. They also have—as sexes—different reproductive functions. Further, their lives are so intimately and inextricably intertwined with men's that, at least at one time, the concept of prejudice against women seemed "preposterous." If not ridiculous, discrimination toward women has often been considered "ambivalent," as if stereotypes are not capable of being positive even if the group is, in general, marginalized. No stereotype is a good stereotype because 1) stereotypes interfere with an individual's ability to act freely, 2) stereotypes, even "good" ones, often lend themselves to corollary negative ones and 3) the so-called "positive stereotype" is generally coupled with negative, latent or unspoken connotation. For instance, Black people are sometimes considered to be stronger and more athletic and, perhaps, thereby, through a terrible, untested warping of logic, considered more tolerant of physical pain (e.g., Mende-Siedecki et al., 2019). When experimenters attributed superior mathematical skills to male, Asian participants, the participants were more likely to infer, latent, unspoken negative stereotypes (Siy & Cheryan, 2016). Furthermore, we speculate that people simply have difficulty imagining that there can be discrimination against white people, even though the power between the genders—even for white people— is squarely with men.

Women have long contended with a profound version of these particular dynamics through so-called "benevolent" sexism, which some once regarded as reflecting this "ambivalence" (e.g., Glick & Fiske, 2001). In addition, the concept that the sheer act of

maintaining close relationships between people from a dominant group and those of a subjugated group as somehow mitigating the potential for prejudice has been justly discarded as absurd. To summarize, having some positive feelings or associations with members of even the marginalized group, itself, is not incongruent with overall bigotry towards that group; in fact, it's inevitable. Black people are sometimes regarded as better athletes, Jewish people as shrewd businesspeople (e.g., Meyer, 1989).

Thus, the fact that there are arguably some "positive" attitudes or stereotypes towards women as a group in addition to negative is obvious (e.g., Eagly & Mladinic, 1994). It is ultimately an argument for the extent of a woman's plight not the mitigation of it. We also suspect that women, in particular, are subject to this kind of fallacy: people may believe that women—especially white women, by virtue of their association with white men—essentially benefit from discrimination. This could lead people to underestimate the discrimination they experience. White women undeniably benefit greatly from being white, just as black men benefit greatly from being male. Further, women may be considered gentler and for instance, at one time, more moral, but do these attributes serve to bolster or hinder them as individuals? Even if such stereotypes were true, do they nullify the fact that women are considered less competent by almost every measure? With increasing momentum, research has begun to demonstrate that women experience great discrimination, especially in STEM (e.g., Van Veelen et al., 2019) and prejudice, such as the deeply-held assumption among women (e.g., Upson & Friedman, 2012) and girls (e.g., Bian et al., 2017) that "brilliance" is the exclusive purview of boys and men as is access to "quantitative" fields and really an arena in which there is potential for recognition or compensation**.** Consider, for instance, the difference between the "home cook" and the glorified "chef." Yet, our studies seek to capture some insight into how misogyny functions on a

generalized construal level—how it targets and harms women in ways that both transcend and support such concrete instances of bias. How do people react to and internalize misogyny that occurs in daily, naturalistic conversation and other interpersonal interaction? Do they even recognize it as misogyny? Do they truly believe that misogyny even *occurs* with any frequency? Few studies have pursued these questions.

Defining Misogyny

Here, we conceive of misogyny as the systematic oppression of women in that they are subject to bias, prejudice, and discrimination—just as one might refer to Black racism or anti-Semitism, for instance. Importantly, we consider its opposite not to be feminism but egalitarianism, as with any other marginalized group. Too often those who advocate for women's rights are labelled feminists, which is problematic both because there is much disagreement concerning what the designation even entails and it also implies that women must *justify* their desire for equality. No other group has that burden. It's implicit to a person's dignity and even personhood that *he* should want the same rights as his fellow man. While each of these forms of bigotry named above encompass their own history, particular struggles, and nuances, they are united in that they are acknowledged to be psychological minorities. That is, they experience systematic subjugation and power dynamics perpetuated by the dominant group—in the case of misogyny, men, or collectively, the patriarchy. They may employ a particular parlance to elucidate an idea or issue, but this tends to serve the wider vernacular—e.g., racism—rather than compete with it.

In contrast, misogyny arguably suffers from a preponderance of terms and associated constructs. For instance, definitions of sexism—which, within this article we will treat

synonymously with misogyny—and especially "feminism" all take on widely different meanings. Scholars and lay people have long proposed various definitions of misogyny. Further, the term which may have attained increased overall prominence in the advent of the #MeToo movement and the 2016 election if google trends and searches provide accurate information (https://www.googletrends.com). People offer varying formulations—particularly between sexism and misogyny. Kate Manne (2017), for one, essentially argues that "sexism" is the behavioral and ideological content of the patriarchal order, and misogyny is the power enforcer of that order. However— "misogyny", "women's rights", the "Woman Question" (to revert to 19th-century debate)—may not benefit from this profusion of names and terms. Racism as a concept makes no such distinctions, yet people are able both to grasp and discuss its intricate trials, tribulations, and history without resorting to endlessly propagating, frequently competing frameworks.

"Feminism," in particular, has been appropriated by so many groups—for and against the advancement of women to equal status with men. In fact, it has taken on so many "waves" and meanings that it is difficult to situate within any purposeful discussion. Crucially, this debate is not merely academic or semantic: we suggest that the diffusion of terms is both a symptom and contributor to the diffusion—and confusion—of overarching problems for women and their daily experience of discrimination. Critically, such circumstances make it difficult to identify and quantify feminism experimentally. Perhaps most revealing is the relative lack of use of "egalitarianism" in discussion of misogyny—because it neglects the severity and pervasiveness of the yoke under which women suffer.

Further, the use and power of feminism subtly suggests that the issues involved are often not as serious as those experienced by other psychological minorities. It also leads to conflicts

among activists who navigate a labyrinth of terminology that often fails to connect one perspective with another in productive dialogue. For instance—as noted above—the construct of black "racism" likewise encompasses these essential elements of a psychological minority: prejudice, discrimination, and unequal power dynamics. Indeed, it could not exist without them, especially the latter. It also employs much terminology of its own—say, microaggressions. However, these concepts *serve* understanding of racism rather than undermine it. There's one single word—racism—and it is potent and powerful in that pointedness. "Racism" is sharp and cuts, where "misogyny," "sexism," and "feminism" can be relatively dull and unformidable. People can and do discuss racism in great depth and nuance without recourse to myriad conflicting constructs and reinvention of terms. They also discuss it far more often if Google Trends and search terms are any indication. In a Google comparison of how often misogyny (and sexism) versus racism are raised as a topic of conversation, racism utterly annihilates the other two (www.googletrends.com). Further the worst slurs of racism cannot even be named in discussion, while misogynistic terms have no such barrier. Importantly, there are terms to describe and discuss the challenges of racism as well. The material point is that they do not tend to conflict and confuse broader, overarching terms like "racism" in the same manner that "misogyny" and "feminism," for instance, are continually reappropriated and reinvented.

To be clear, a simple dichotomy between "misogyny" and "egalitarianism" does not at all indicate a simplicity of issues. Nor does it suggest that various intersectional —overlapping identities among groups and individual—aren't real and relevant. As already stated in referring to racism and anti-Semitism, different minorities contain their own content and specific struggles, requiring necessary terminology (Shih et al., 1999). Feminism, though conjured into discussion by Charles Fourier nearly two centuries ago in 1837 (Grogan, 1992) when the sheer concept of

equality between the sexes was almost fantastical—and gender as a separate concept, unthought of (by men)—perhaps served a purpose and may even have long served one. It is worth noting, however, that voices like Mary Wollstonecraft were already arguing for "equal rights" decades before the inception of this word, feminism. Regardless, we suggest, misogyny has graduated from that day and must and deserves to be treated with the dire import of a full-fledged form of bigotry.

The Stamina and Pervasiveness of Misogyny

One fact seems glaring: as mentioned, the sheer continued use of "feminism," however defined, implies an obligation to justify misogyny as a real-life phenomenon. It manifests in a word the need for a defense of what is implicit in the relations among other groups: the recognized desire for equal rights and freedom that is intrinsic to the human condition, reluctant as the dominant group may be to grant those rights. No one wants to live in in a cage: among men, this is well-accepted. In consideration of women, it's not a question of whether there is a cage, it's how restrictive this cage may be. At best, the questions become how flexible these restrictions are and how cleverly composed of "benevolence" and indeed vastly differing definitions of even sexism (e.g., Umera, 2012). At worst, misogynistic practices are among the most sinister and cruel forces on Earth, perpetuating prejudice, control, and negative attitudes in many domains—and also dominance through institutional violence in some cultures **(**Fiske & Rai, 2014). Repugnant if not as harrowing, women are frequently told, especially in Western countries that they, so to speak, like their cages. For example, people—men and even women—may regard what has been euphemistically termed "benevolent" sexism—ranging from the silliness of holding a door open to added protection in a particular situation to the wearing of a

hijab— as a something of a tacit agreement or contract (e.g., Glick & Fiske, 2018). Again, it's original use by Glick was in demonstrating how "benevolent sexism" was also negative and used by men to manipulate and reward women for conforming to sexist norms. However, as a consequence of this confused terminology, people (men) may believe in an apparent acceptance that justifies spurious arguments that women are different from other minorities and should accede to other forms of bias and domination. People—men, women, and those of other gender identification—alike value their autonomy and agency. No other minority discusses "choice;" if nothing else, choice is inherent to freedom.

To date—as briefly noted above—there has been accumulating research, that women endure discrimination on many levels and in many domains, even limiting discussion to American society, alone. As mentioned, women and girls as young as six-years-old believe that only boys and men are capable of attaining brilliance (Bian et al., 2017). Studies also indicate that women believe they are relatively poor at quantitative and spatial tasks (e.g., Shih et al., 1999), all of which are likely false (e.g., Hyde, 2005). Nonetheless, these so-called facts without a doubt engender false beliefs which then contribute to many cascading effects on women's choice of field and career (e.g. Alon et al., 2015), their self-esteem and self-efficacy (e.g., Dickerson & Taylor, 2000), their aspirations, and likely their interactions with others. One series of experiments demonstrated that men considered themselves to be categorically smarter than women at all levels of intelligence, rather than just the extremes (Cooper et al., 2017). Collectively, studies have come to illustrate that these so-called differences are illusions and the product of socialization and women's inferior status. For instance, Bian and colleagues (2018) found that, among other results, women showed less interest and evaluated internship descriptions emphasizing intellect as incongruent with their abilities, registering anxiety and fear

of belonging. Importantly, these effects also were not limited to STEM jobs—which tend to receive more attention—and also applied to humanities and unspecified subject areas. Further, studies have shown that—as with other marginalized groups—women exhibit stereotype threat, as noted above. Women have and do suffer terrible physical abuse, both at the hands of close others, and, in some countries, systemically. Such a state of vulnerability inevitably strengthens stereotypes about women's weakness.

Such an overview is just a sampling of the burgeoning research demonstrating that women are a true psychological minority and that they suffer greatly from their inferior status. Many theories have presented themselves for why misogyny has remained so stalwart. Eagly's social-role theory (e.g., 2000) has long attributed widespread stereotypes to a fundamental division of labor, with women as "homemakers" and men as "paid workers." Such differentiation has probably contributed to discrimination, but research suggests that so-called "separation of spheres" are not the only causes. Indeed, it is dangerous to presume ultimate causal assumptions, even among what seem logical, well-accepted evolutionary likelihoods. For instance, Anderson and colleagues (2023) startled even themselves with the discovery of evidence that challenged the widely accepted "males as hunters, females as gatherers" early-human paradigm. Their review of archeological findings—e.g., women buried with weapons—have suggested not only that women hunted but that at least during the Holocene era some 13,000 years ago, this was common practice across cultures around the world.

Perpetually "other" yet not quite "out-group" in the sense that it is widely used in the common canon, women thus occupy a unique position in society for exploring moral fault-lines, revealing interesting nuance about categorization in general. This juxtaposition is another way in which the current studies contribute to how people fundamentally think about people as

individuals and groups. Misogyny likely remains stubbornly prevalent because its origins are ancient, the dominant group (men) remains very powerful, and men and women's lives are inextricably intertwined, which can create the misconception that misogyny as described cannot occur on a structural level. For this reason, women—perpetually lesser—are perhaps a convenient target of vitriol for insecure men (e.g., Cross et al., 2019). Further, it is often associated with such unsubstantiated differences in intellectual proclivity and essentialism. Finally, there are, of course, the juggernaut findings: 1) People connect gender with the essence of being human (e.g., Martin & Mason, 2022) and 2) Men are considered the prototype for even being human at all. Consistent with these findings, Pietraszewski and colleagues **(**2023**)** found in a study exploring alliances that people—under particular circumstances— could easily align with those of another race but not of another gender. Finally, Momennejad and colleagues (2019**)** presented simulations that argued strongly for how detrimental sexism is for men, women, and institutions—which, when comparing statistics, is not readily apparent.

Lacking the consistent vernacular to articulate these problems in a nuanced yet succinct and identifiable way, it seems all the more likely the misogyny would be underestimated. Thus, exploring the nature of misogyny as it occurs in the everyday not only contributes to understanding of the prejudice against women per se. It also provides an almost unique opportunity of investigating how categorization functions and bigotry can subsist through within a network of one's closest relations: fathers, brothers, husbands, sons, friends. In short, we suspect that people don't truly conceive of women—the full category, encompassing all races and creeds—as a true psychological minority unto itself: Women don't suffer the prejudice, discrimination, or indignity of a true minority.

The Current Studies & Hypotheses

In the following experiments, we set out to examine perceptions of misogyny by comparing it to Black racism. We know of no other experiment that has explored misogyny through such a method. Our objective was to capture snapshots of bigotry as they can and do occur frequently and fluidly in dyadic and group dynamics like those that might take place in an office setting. Crucially, we tried to strip the language we used in each of our between-subjects vignettes so that they included minimal information other than the fact that an act of bigotry had occurred against either a white woman (the misogyny condition), a Black man (the racist condition), and a Black woman (misogyny + racism, hybrid condition), with a neutral altercation with a white man (control condition). Our primary hypotheses were the following. Dependent variables are in Table 2:

TABLE 2.1: Chapter 2 hypotheses

Hypothesis 1; "Bigotry Light:" Racism is considered substantially more offensive and meaningful than misogyny, which will be reflected in participant evaluations of slurs and insults.

- **Prediction:** We predicted that participants would consider racist slurs as more indicative of the protagonist's essential nature in comparison to misogynistic slurs. As discussed, we suspected that, for myriad reasons, this might occur because people tend to underestimate the frequency and power of misogyny.
- More specifically with regard to our dependent measures, we theorized that people would rate the Agent in the "racist" condition to be much "worse"—e.g., more racist, more bigoted, less moral, more deserving of punishment, etc.—than the other conditions. We expected that racism would be followed by the misogynistic/racist, hybrid condition (from here termed the hybrid condition), and then the misogynistic condition, which we thought would likely differ significantly from the neutral condition. With the exception of the misogyny

TABLE 2.1 (continued):

question (discussed below), we expected that this basic pattern would persist through all of our DVs to varying degrees.

Hypothesis 2; **"*Two Wrongs* Don't Even Make One Full Wrong:"** Misogyny in combination with racism will be rated as less offensive and wrong than racism alone.

- **Prediction:** Interaction-- Demonstrations of racism and misogyny in the form of slurs that apply to Black women will fall in between, with misogyny diluting the effect of "pure" racism rather than compounding it. Instead of producing an additive effect as might be supposed, we expected that focus on misogyny will—somewhat illogically—impress people with overall lesser force. People may presume that Black women are somewhat shielded from the worst aspects of racism because, as women, they have greater need and are provided more protection. As stated above, we also predict that people simply have more difficult believing and imagining that women—in general, but especially white women—suffer prejudice and are a psychological minority. The misogyny component of being a black, marginalized person has a subduing effect. Compared to the other two slur conditions, we predict, therefore, that the transgressions against Black women will be rated as less "wrong" than the same behavior in racism. In addition—and for the inverse reason, we expected that the hybrid condition would most likely be considered worse than the misogyny condition.
- Prediction: As discussed above, we expect that for each DV, except sexism, that the racism condition would be judged "worse" than the hybrid condition. We tentatively predict it will tend to fall between racism and misogyny.

Hypothesis 3 ***(experimental);*** **Gateway Hypothesis:** We hypothesized that misogyny can sometimes function as a kind of negative conduit or "gateway" into other forms of bigotry or radicalization.

TABLE 2.1 (continued):

- **Prediction:** If this hypothesis were to be true and, moreover, reflected in the current experiment, we predicted that people who read the version of the scenario in which the white man (Agent) uses racist slurs will be evaluated as more racist and also, crucially, considered sexist. To be clear, there is no information in that vignette that provides any information about sexism; thus, if participants report that the Agent is sexist, they are inferring that misogynistic tendencies from the fact that he has exhibited racist behavior. We further expect that the inverse will be less robust—if it exists at all—for the sexist DV. As briefly noted above, research has provided some reason to believe that men bond over denigrating women. Since men's "alliance system" does not identify women as potential candidates for coalitions and instead regards them almost inescapably as "other" (Pietraszewski et al., 2014), women may instead be candidates for *targets* or at least the root cause of such affiliations. In other words, if men can bond across races through commonalities—including common "enemies"— the fact of being a woman may present a contender for this "commonality." When a man—particularly a member of a minority group who has experienced some act of racism, say—goes home, he has an eligible member of a group on which to vent his frustration: his wife, his daughter, his mother (e.g., Cross et al., 2019).
- As explained above, the "sexist" and "racist" DVs together provide the test of this hypothesis. As noted, we hypothesize that men—especially those prone to bigotry—might bond over ridiculing women as a group that is, by some standards, always "second class" and "other" but crucially not an "outgroup." Further, we theorize that misogyny would be something like "bigotry light," suggesting poor but not quite offensive behavior. For the reasons, we predicted that people would infer that men who displayed prejudice or discrimination against other groups were also likely to be misogynistic. We expected that the hybrid condition would be rated as most sexist, followed by the misogynistic condition. However, we also

TABLE 2.1 (continued)

predicted that the Agent in the racist condition would also be judged to be misogynistic (as measured by differing from control), despite that particular

- vignette alluding to nothing of the kind. To reiterate, we did not expect, in this case, that the hybrid condition would be diluted by the compounded forms of bigotry. On the contrary, we thought it likely that the racist component would likely boost the hybrid condition above its sister condition—with a white woman as "target." Importantly, in the racist measure, we did not expect the misogyny condition to differ from control. In other words, we did not expect the inverse: for people to infer racism from misogynistic information.

OPEN SCIENCE

Both studies were pre-registered on The Open Science Network (OSF), which is maintained by The Center for Open Science (COS), a non-profit organization dedicated to the integrity and transparency of scientific methods and findings. For each experiment, we recruited participants using Cloud Research, an online platform created and supported by Amazon. We limited eligibility to people ages 18 years and over, living in in the United States, and holding a high standing within Cloud Research—meaning past work indicated that they were attentive and completed tasks in a timely manner. Otherwise, we did not target any group or demographic or ultimately make any exclusions from the collected data.

STUDY 1

Methods

Participants. 189 participants completed the experiment in a median time of 6.4 minutes and were compensated $1.50 for their time. They were recruited from Amazon's Cloud Research.

We had planned for 50 participants/cell, but due to drop-out we ended up with fewer in some cells. Power analyses have demonstrated that this sample size is more than adequate for extracting our intended effects. No participants were excluded from analyses. The sample had a mean age of *M*=39.2 (*SD*=10.7); 116 were female.

Design & Procedure.

Study 1 (as well as Study 2) employed a 1(Agent/James: white man who loses his temper) x 4(Target/Taylor: White woman—misogyny condition vs. Black man—racist condition vs. Black woman—racist + misogyny condition vs. White man—control condition) between-subjects design. Participants were randomly assigned to one of four resulting conditions, which each contained approximately 50/cell. Participants then read a brief vignette in which a "target," always named Taylor, was openly critical of the "Agent," who was always named "James." The scenario then describes how James reacts by losing his temper and calling "Taylor" pejorative terms. Importantly, in the three "bigotry" conditions: the misogyny, racism, and misogyny/racism (from here, hybrid) conditions, the terms were described as unspecified slurs or in the case of the control, "insults". For instance, in the racist condition, Taylor is identified as a Black man, and the slurs designated as those "applied only to Black people." The exact wording for all conditions is presented below in Table 2:

TABLE 2.2: Description of the manipulation in Study 1. James, the offender who uses vicious language is always a White man and is bolded, as is the control version of Taylor; Taylor, the target, is a different demographic in each condition and is color-coded. These font deviations were not used in the presentation to participants.

All conditions first began with the following phrasing. The two misogyny conditions used female pronouns, of course, while the other two conditions used male pronouns when referring to Taylor.

TABLE 2.2 (continued)

You work in the same department as two people, **James** and Taylor. Taylor is known to deliver tough, blunt criticism and by general reputation does not suffer fools. Most people know that he (or she) doesn't particularly respect **James**, whom he (or she) regards as a

From here, the vignettes diverged according to condition:

MISOGYNY CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses **his** ideas as unproductive and stupid. In a moment of aggravation, **James**, **a White man,** loses his temper and viciously insults **Taylor, a white woman**, using slurs applied only to **women. He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

RACIST CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses **his** ideas as unproductive and stupid. In a moment of aggravation, **James, a White man,** loses his temper and viciously insults **Taylor, a Black man**, using slurs applied only to **Black people**. **He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

MISOGYNY & RACIST CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses his ideas as unproductive and stupid. In a moment of aggravation, **James, a White man,** loses his temper and viciously insults **Taylor, a Black woman,** using slurs applied only to **Black women. He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

CONTROL CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses his ideas as unproductive and stupid. In a moment of aggravation, **James,** a **White man**, loses his temper and viciously insults **Taylor,** who is **also a white man,** calling him names in the course of it. **He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

At this point, each participant answered ten questions about James—to whom we will refer as the "Agent"—which served as our dependent variables. All ten questions were

randomized and presented in counterbalanced order. While we anticipated a high degree of correlation among at least some of the measures, the questions were intended to capture different manifestations of how an observer might interpret behavior—i.e., whether it reflected bigotry and if so, what the particular content of bigotry would be—and more, specifically, how it would relate to our three primary hypotheses. In addition, we expected that questions essentially calling for *characterizations* of the Agent might differ in pattern or intensity than those that focused more on his *behavior* in the respective, one-time experimental incident. We expected that that the pattern would be preserved but that the effect size might be greater because people would be more likely to commit to condemnations of behavior rather than overall characterizations.

All 10 dependent variables referred to a scale that visibly ranged from 0 to 100, with 10-integer demarcations. However, participants could record their responses on an integer level. We chose this scale to provide more granularity to participant choices so that they had the opportunity to choose with greater specificity than is usually accorded by a 7-point or 9-point Likert scale. Each scale also was marked with written guides at both extremes—e.g. "Extremely Racist" at one end, "Extremely Egalitarian" at the other, and "Neither" or "Unsure" at the midpoint. The participant registered their decision using a cursor that rested at the midpoint as a starting point. All DVs were randomized and counterbalanced. The exact wording of the questions is printed below in Table 2:

Dependent Measures

Below we list all DVs included. Those in navy blue charged participants with making characterizations about the Agent. Those in green targeted appraisals of his one-time *behavior* in the incident that participants read about. We also speculated that people might be wary of having

associations with the target, which would be a practical consequence of the Agent's bigotry (printed in orange below).

We used the Holm-Bonferroni method (Holm, 1979) to control for family-wise error, which we set at 0.05. We first applied it to the 10 DVs. Dividing by successively fewer numbers of tests—or more conservatively, using Bonferroni--made virtually no difference as each omnibus F-test for the respective DV was much below 0.001. The one exception was for the "Aggression," DV which had a p-value of almost 0.001, and by either correction method remained clearly significant. In terms of correcting for multiple comparisons for the post-hoc tests for each DV, we also used Holm Bonferroni, ranking the p-values and comparing them to the adjusted value.

We expected that the DVs would coincide with the general predictions outlined above. The following Table 2.3 provides the exact wording of all DVs:

TABLE 2.3:

All Dependent Measures: dark blue represented questions calling for characterizations, green for questions evaluations of behavior, and yellow for a question that is personal to the participant

1) Based on his behavior, is Agent [James] likely a bigot?
2) Based on his behavior, is Agent likely racist?
3) Based on his behavior, is. Agent likely sexist?
4) Agree or Disagree: Agent is probably a moral person overall.
5) Was Agent's behavior good or bad?
6) Should Agent be punished or praised for his behavior?
7) Is Agent a jerk?
8)Do you think you would be friends with James?
9) Agree or Disagree: Agent's actions were aggressive.
10) Agree or Disagree: James is probably a closed-minded person who doesn't respect people who are different from him.

After reading the scenario and answering the questions, participants were prompted to provide some demographic information, thanked for their participation, and paid.

Analyses for Study 1 & 2: With a few additional exceptions, we regressed the four conditions on the each respective DV and ran a one-way ANOVA. We also used the Holm-Bonferroni correction to account for the 10 DVs, setting the family-wise error at 0.05. In addition, we used the same method when conducting post-hoc tests for each individual DV.

Study 1: Results

We found substantial support for all three hypotheses: "Bigotry-Light," "Two Wrongs," and "Gateway." Once again, we used the Holm- Bonferroni method for multiple comparison. With regard to the question of correcting for each of the 10 DVs, it was not necessary to apply Holm Bonferroni, as each of the F tests was well below even a stringent Bonferroni correction 0.05/10=0.005, as is evident in Table 1. A complete list of the descriptive results and statistics are available in Supplementary Materials.

Hypothesis 1; "Bigotry Light"*:* Racism is more offensive and meaningful than misogyny.

Hypothesis 2; "Two Wrongs"*:* Rather than compounding judgments of racism, the hybrid condition (racism + misogyny) will be rated as less offensive than racism, alone, and will likely fall somewhere between the two other experimental conditions (pure racism and pure misogyny).

Dependent Variable: Bigotry

As depicted in Figure 2.1A, a one-way ANOVA yielded an effect of bigotry, $F(3, 186)=24.4$, $p<0.001$, $\eta^2=0.282$ and an adjusted coefficient of determination, $R^2=0.271$. As predicted, the Agent in the racism condition was rated as most bigoted, ($M=69.9$, $SD=18.6$, 95%CI=[64.5 -75.3]), judged to be higher than the misogyny condition ($M=45.6$, $SD=30.5$, 95%CI=[48.3-54.3]; $t(102)=4.45$, $p<0.001$, $d=0.892$) but only marginally higher than the hybrid condition, ($M=60.8$, $SD=30.3$, 95%CI=[52.4-69.2]; $t(82.6)=1.79$, $p=0.078$, $d=0.358$)**.** Pair-wise comparison of the two misogyny conditions was significant, ($t(94.6)=2.50$, $p=0.015$, $d=0.501$)**,** with the Agent in the hybrid condition rated as significantly more bigoted**.** The difference

between the misogyny and the control conditions, (M=27.3; SD=21.8, 95%CI=[6.18-33.5]) was also significant: t(83.3)=3.36, p=0.001, d=0.069.

These results provided strong evidence for the first two hypotheses. In accordance with the Bigotry-Light Hypothesis, racism is considered much more offensive than misogyny, based on a robust effect size. In striking conformity with the Two-Wrongs Hypothesis, not only was the effect of being targeted as female and Black not additive, but the combination was sub-additive, its rating below that of the racist condition, though the difference was only marginal. The misogyny component seemed to severely dilute the impact of racism on its own. Further, these effects occurred in a most stark and direct question about whether the respective Agents had experienced bigotry. Each of the vignettes (with the exception of control) stipulated that they had in fact, experienced bigotry—without of course stating what words were used. Though the difference between racism and the hybrid condition was not significant here, it was significant in other DVs, like morality (Figure 2.1B below). It was also different from misogyny for the majority of DVs. Further, while all descending comparisons were not always significant, the same pattern prevailed with each DV, often with strong effects: Racism was judged worst, most deserving of punishment, least likely to invite friendship, most aggressive, etc. with hybrid rated next highest—or lowest as the case may be—followed by the misogyny and finally, the control condition. There was one exception—relating to our third, Gateway Hypothesis. We present those findings in discussion of the DVs that ask if the Agent is "misogynistic" and "racist," respectively.

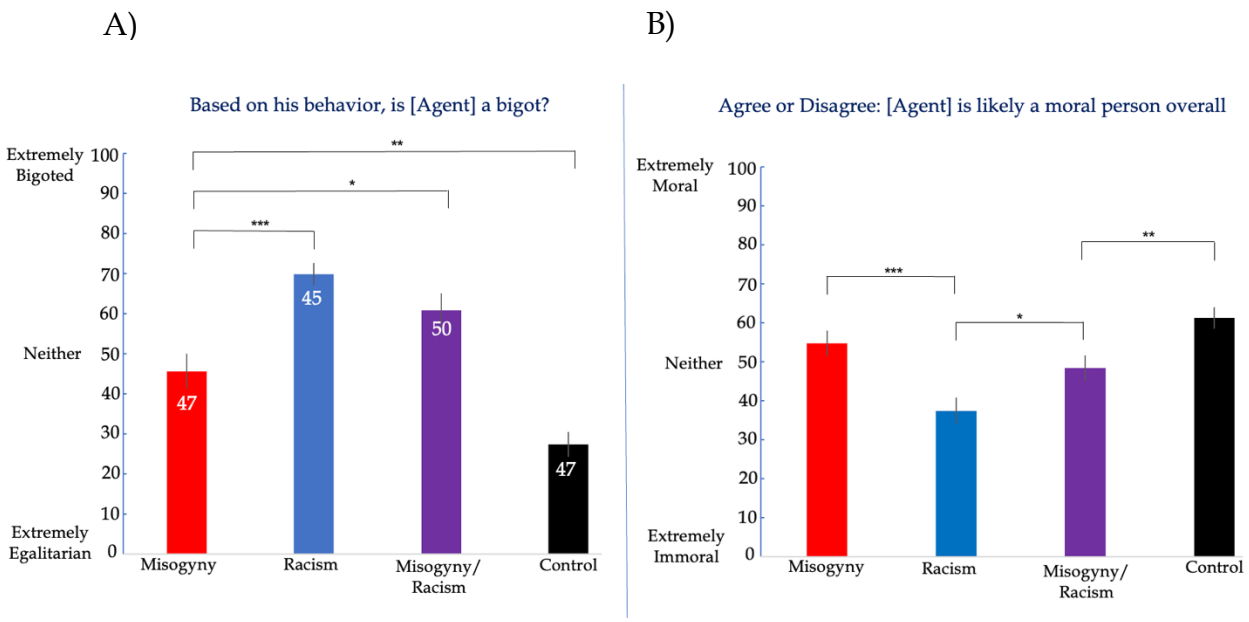

Figure 2.1. Evaluations of A) characteristic bigotry of the agent; B) and characteristic morality of the Agent. Titles are the respective questions presented to the participants, and bars represent SEs. As represented in Figure 2.1A, n=47 for the misogyny condition, n=45 for the racist condition, n=50 for the hybrid condition, and n=47 for the control condition.

Dependent Variable: Morality

As likewise pictured in Figure 2.1B, "morality" displayed a similar pattern to "bigotry," with a main effect of morality: $F(3,186)=9.80$, $p<0.001$, $\eta^2=0.136$; R^2(adjusted)=0.122. In this case, racism (M=37.3, SD=22.6, 95%CI=[30.7-43.9]) differed from all conditions. Interestingly, none of the other comparisons were significant, except for that between hybrid and control. The behavior of the Agent from the racist condition was rated significantly less moral than the Agent in the misogyny condition, (M=54.7, SD=22.3, 95%CI=[48.3-61.1]); $t(89.7)=3.71$, $p<0.001$, d=0.774) and the hybrid condition, (M=48.2, SD=24.0, 95%CI=[41.5-54.9] ; $t(92.8)=2.27$, p=0.025, d=0.466). As mentioned, the hybrid condition was also higher than the control(M=61.1, SD=19.1, 95%CI=[55.8-66.5]; $t(92.8)=2.96$, p=0.004, d=0.594). Once again, misogyny was not different from control: $t(95.0)=1.38$, p=0.172, d=0.312.

These results provided particular support for both the Bigotry-Light and Two-Wrongs Hypotheses. Once again, the racism condition was overwhelmingly dominant, while the only

other comparison that differed was hybrid from control. Especially given the nature of the question—morality—the fact that misogyny and control did not differ is striking. One might expect that any form of bigotry would be considered less moral than uncharged insults.

Dependent Variable: Goodness of Action

Analyses produced a main effect of "goodness of action:" $F(3,186)=12.2$, $p<0.001$ $\eta^2= 0.164$, with an adjusted $R^2=0.151$ for the model. The racism condition, (M=13.3, SD=15.2, 95%CI=[8.86-17.7] fell below both misogyny (M=21.9, SD=20.0, 95%CI=[16.1-27.7]; $t(85.6)= 2.31$, p=0.023, d=0.481) but only marginally differed from hybrid, (M=20.5, SD=22.2, 95%CI=[14.0-26.6]; $t(87.1)= 1.84$, p=0.069, d=0.371. Yet again, the misogyny conditions did not differ, though both were rated below control (M=48.0, SD=17.7, 95%CI=[43.0-53.0]). Pairwise comparison between hybrid and control yielded $t(93.0)= 3.95$, $p<0.001$, d=0.838. The difference between misogyny and control was significant at $t(91.1)= 2.51$, $p<0.001$, d=0. 0.771.

These findings also supply especially strong support for both the Bigotry-Light and Two-Wrongs Hypotheses. Racism was, once again, lower than all other conditions. The hybrid condition was almost even with misogyny, suggesting that gender utterly eclipsed race in the assessment of this condition. These results are also especially notable because the question was a straightforward assessment of whether the behavior was good or bad. One might presume that it would require very little cognitive effort to assess the experimental conditions: each quite bad and the hybrid condition especially so, encompassing racism and misogyny, as it does. In comparison, for instance, how likely the participant would be to "befriend" the Agent or how much the Agent deserved punishment might be questions that would require more reflection that might lead people to call on stereotypes "favoring" racism over misogyny. Further, as predicted,

at least in this case, judgements of behavior appear to be stronger than judgements of overall character.

A) B)

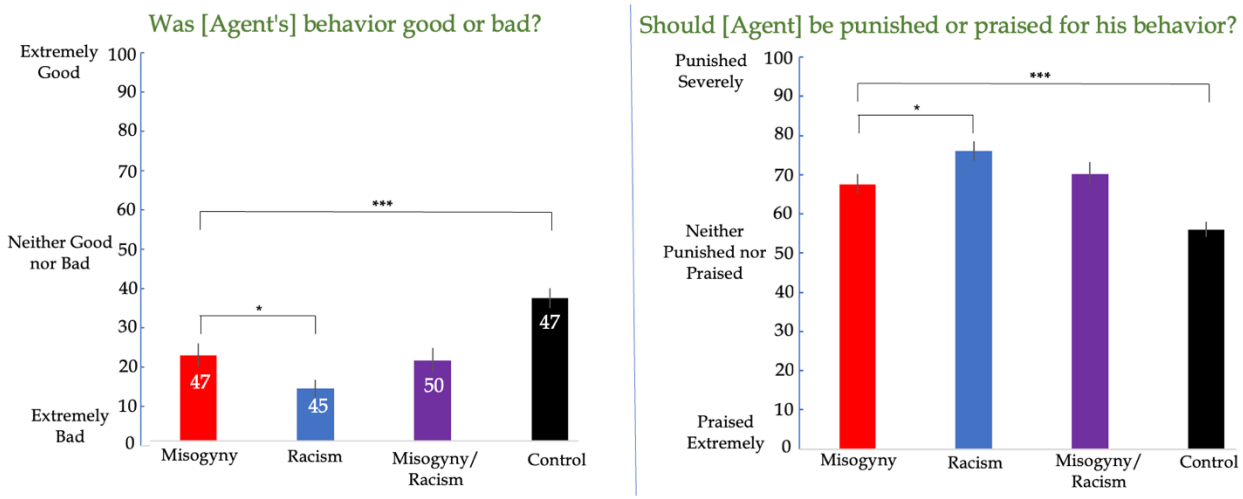

Figure 2.2. Evaluations of A) how good or bad the Agent's actions were; B) and judgments of whether the Agent should be punished. Titles are the respective questions presented to the participants, and bars represent SEs. As represented in Figure 2.2A, n=47 for the misogyny condition, n=45 for the racist condition, n=50 for the hybrid condition, and n=47 for the control condition.

Dependent Variable: Punished or Praised

As pictured in Figure 2.2B above, analyses once again produced the same pattern as all other DVs (except for sexism, discussed later), with a main effect of punishment, $F(3,186)=10.6$, $p<0.001$, $\eta^2=0.146$ and adjusted $R^2=0.133$. The Agent in the racist condition, (M=76.0, SD=16, 95%CI=[71.3-80.7]), was rated highest in the direction of punishment and was significantly greater than the corresponding Agent in the misogyny condition, (M=67.6, SD=17.6, 95%CI=[62.6-72.6]; $t(89.9)=2.38$, $p=0.020$, $d=0.393$), though not the hybrid condition: M=70.1, SD=21.7, 95%CI=[64.1-76.1];]; $t(90.0)=1.51$, $p=0.136$, $d=0.231$. Pair-wise comparison between the misogynistic conditions was not significant, with the hybrid condition only barely exceeding misogyny and rated as more deserving of punishment: $t(93.0)=0.620$, $p=0.536$, $d=0.392$. All experimental conditions differed from control, (M=56.0, SD=13.7,

95%CI=[52.1-59.9]) with large effect sizes. Control and misogyny produced the following: $t(86.8)=3.56$, $p<0.001$, $d=1.17$. This DV represented one in which participants clearly differentiated bigoted behavior from behavior that was merely aggressive or intense, as demonstrated by the large differences between control and the experimental conditions.

Hypothesis 3; "The Gateway:" Theorizing that sexism will be regarded as less pernicious and even a possible basis for male bonding, we speculated that people would infer misogynistic characteristics even when supplied with only racist information. We predicted, therefore, that the racist condition might be rated as sexist, though no such information has been provided in that scenario. We also predicted that the inverse would not be true: people would not infer racism from misogynistic information. If they were to make such an appraisal, we expected that this opposite inference would not be nearly as strong.

Dependent Variable: Sexism

As pictured below in Figure 2.3A, sexism was the one dependent variable that did not display the same basic pattern of results as the others. There was a main effect of sexism: $F(3,186)=12.1$, $p<0.001$, $\eta^2=0.163$; R^2(adjusted)=0.150. In striking support of the Gateway Hypothesis, all experimental conditions differed from control, ($M=26.6$, $SD=23.2$, 95%CI=[20.1-33.1]), and none from each other. The three were almost even, as depicted below in Figure 2.3A. With due consideration that this is could be construed as a null finding, it nonetheless appears meaningful, given that all conditions differed from control, for one. Further, it fit our hypothesis—perhaps more extremely than predicted. As there was so evidently no difference among the experimental conditions, we provide each pair-wise comparison with

control. The racism condition vs. control condition produced the following: (M=50.3, SD=22.1,95%CI=[43.8-56.7], t(91.0)=5.05, p<0.001, d=1.05). Misogyny, (M=55.2, SD=31.3, 95%CI=[46.3-64.1] was significantly higher than control at t(85.0)=5.07, p<0.001, d=1.043). Finally, the hybrid condition, (M=53.8, SD=24.0, 95%CI=[45.7-61.9]) was much greater than the control, (t(92.9)=5.13, d=1.03, as likewise reflected in the other respective effect sizes.

Even given our prediction, these results provided strong support for our "Gateway Hypothesis" and also our first two: Bigotry-Light and Two-Wrongs. Two particular aspects are especially prominent. None of the conditions—even misogyny and hybrid—were rated very highly, barely above the midpoint; indeed, the misogyny condition (the highest of the three) did not differ from the midpoint: t(46)=1.14, p=0.261. Secondly, the mean for the racism condition was almost exactly the same as the other two. Racism, which logically should rest somewhere near control, produced a very large effect size in comparison to control that rivaled those of the misogyny and hybrid conditions, respectively. Further, in spite of our predictions, the hybrid condition was not higher than the misogyny condition—in magnitude, it was, in fact, lower. These results become all the more glaring when comparing them to the findings from the racist DV.

A) B)

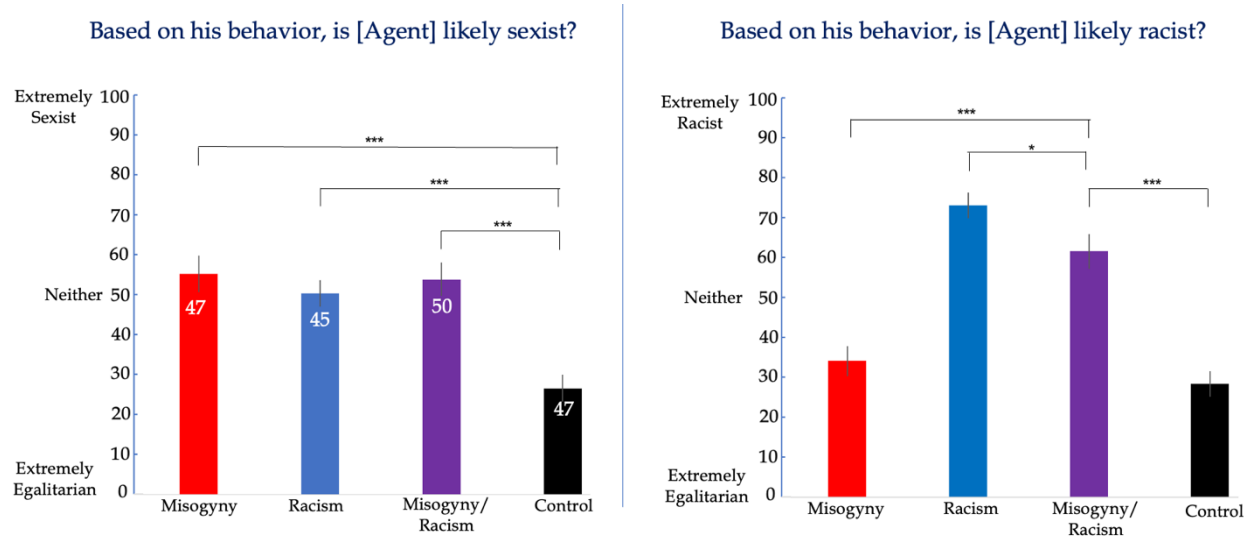

Figure 2.3. Evaluations of A) characteristic sexism of the agent; B) and characteristic racism of the Agent. Titles are the respective questions presented to the participants, and bars represent SEs. As represented in Figure 2.3A, n=47 for the misogyny condition, n=45 for the racist condition, n=50 for the hybrid condition, and n=45 for the control condition.

Dependent Variable: Racism

As reflected In Figure 2.3B, racism displayed the same pattern of results as the other eight DVs (excluding "sexism), with a main effect at $F(3,186)=33.5$, $p<0.001$, R^2(adjusted)=0.340, and $\eta^2= 0.351$. All conditions differed from one another, with the sole exception of misogyny and control, which, logically, did not. Again, applying Holm Bonferroni, the racism condition, (M=72.8, SD=21.4, 95%CI=[66.6-80.0]) was rated higher than the hybrid condition (M=61.2, SD=31.1,95%CI=[52.6-69.8]; $t(87.4)=2.12$, $p=0.037$, $d=0.428$) and much higher than misogyny, (M=33.9, SD=25.2, 95%CI=[26.7-41.1]; $t(88.8)=7.98$, $p<0.001$, $d=1.66$), as is readily apparent by the large effect size of the latter. The Agent in the hybrid condition was also rated far more racist than the misogyny condition: $t(93.0)=4.78$, $p=0.001$, $d=0.964$. As was

equally, misogyny did not differ from control, (M=56.0, SD=13.7, 95%CI=[52.1-59.9]; $t(90.9)=1.18$, $p<0.240$, $d=0.242$)

All conditions differed from the control condition, except for misogyny. This finding contrasts starkly with the results from the misogyny dependent measure, for which even the racist condition was considered to imply sexism. Indeed, such a result—misogyny failing to differ from control-- would not seem so notable, except in the overwhelming contrast to the sexism measure. As already discussed, the racist condition was lower than the other two only in magnitude and very slightly. This racism DV—perhaps in conjunction with the misogyny measure—probably provides the best simultaneous support for all three hypotheses. All conditions—except for misogyny and control— were different in the descending order of the pattern that has permeated all DVs, in accordance with the Bigoty-Light and Two-Wrongs Hypotheses. Further, the average rating of all four condition in the racism DV was $M=48.8$, SD=31.2, which is comparable to the average of all conditions in the context of the sexist DV at $M=46.4$, SD=29.0; yet, the mean for only racist conditions (racism and hybrid) was $M=66.8$, $SD=27.4$, while the two sexist conditions (misogyny and hybrid) came to an a mean of only $M=54.4$, $SD=30.0$, barely above the midpoint. It is possible that people might make inferences the other way—e.g., misogyny may suggest racist proclivities as well—but there was no evidence for it in these two measures. The results produced by the two DVs were wildly different, and the misogyny condition was only greater than control in magnitude, as discussed above.

Dependent variables: Jerk, Friend, Aggression, Respect (All relevant pair-wise tests available in Supplemental Materials)

While we did not include graphs for the measures, each of these DVs produced the same pattern of results, as repeatedly discussed (see Tables A.1 and A.2 in the appendix). In accordance with our suspicions that misogyny may be regarded ambivalently as somewhere between rude and offensive, the "jerk" DV and "friendship" DV potentially represent interesting contrasts. The racism and hybrid conditions differed from control for every DV. Misogyny did, in fact, differ from control for most, but notably, not "morality" and not "friendship." In the context of the "jerk" DV, misogyny was different from control; placed directly in contrast with the "friendship" DV is perhaps revealing. One might presume that a person would not befriend —and be reluctant to admit an association (for fear of social repercussions)—with someone they regard as morally lacking. Yet, they might befriend a person who is a jerk now and then. When considering this interpretation, the morality measure seems striking. Failing to register a moral difference between a woman experiencing an slurs versus a white man confronting some name-calling from another white man seems quite telling.

Otherwise, as stated, the "respect" condition was largely representative of the others, with marginal differences between racism and hybrid and between hybrid and misogyny. As with every other DV—except sexism—racism was significantly different than misogyny with a large effect size. The Agent in the misogyny condition was also apparently considered much less respectful of people who are different from him than the Agent in the control condition. "Aggression" proved to be a slight anomaly. The same pattern held, yet all four conditions were rated as fairly aggressive, the three experimental conditions all above 70. The misogyny condition was not different from control but was high. The racism condition was close to ceiling. Once again, tables detailing the models for each DV and pair-wise comparison for Study 1 are available in the Appendix, as is a table detailing the results for each DV in Study 2.

Study 1 Discussion

Study 1 provided strong evidence for all three of our hypotheses. Each measure—with the sole exception of the "sexism" DV—satisfied the Bigotry-Light Hypothesis that racism would be rated worse than misogyny. As noted, each of those nine DVs also reflected the same pattern—in magnitude— with the racism condition being the least desirable rating, followed by the hybrid condition, the misogyny condition and, finally, the control condition.

Each DV also supported the Two-Wrongs Hypothesis, which predicted that the hybrid combination of racism and misogyny would not be additive. In fact, it never even exceeded racism, even in magnitude. Indeed, the hybrid condition largely conformed to our loose prediction that it would fall somewhere between misogyny and racism. In magnitude, the hybrid condition always occupied the middle position (except for the "sexism" DV). However, the hybrid condition did not always significantly differ from racism or from misogyny for a few critical DVs—namely, the punishment, jerk, friendship, and aggression DVs. Such results lend even greater credence to the Two-Wrongs hypothesis. One might expect that such serious and direct questions—among the most relevant to the respective scenarios— would have revealed some difference between at least misogyny and hybrid. For even stronger evidence of Bigotry-Light, the misogyny condition did not always differ from control for some conditions, including how moral the Agent likely was and whether the participant would be likely to befriend the Agent, as already discussed. Such findings suggest that people find misogyny extremely rude—and even "aggressive"—but perhaps not a stigma (at least on the same level as racism) and not necessarily all that socially undesirable in the long run.

Finally, the misogyny DV—especially in comparison to the racism DV—provided evidence for the Gateway Hypothesis. As discussed above, the racism condition included no

mention at all of anything that could be construed as misogyny, so the data strongly support the interpretation that people inferred misogyny from racist content. In contrast, in the racist DV, misogyny did not even differ from control and was well below the midline toward "egalitarianism." Particularly in combination, both DVs bolster the conclusion that people did, in fact, extract misogynistic tendencies from racist information.

STUDY 2: Replication using explicit language

In Study 1, we deliberately phrased the bigotry described in each of the experimental conditions as very general, leaving people the freedom to imagine what the impact of "slurs" against a "white woman," a "black man," and a "black woman" might experience in real life. We wanted to include as little prompting as possible and to let people's own perspectives on these issues dictate. In this second follow-up study, we were interested in exploring if the simple alteration of explicitly using the words, "sexist" and "racist" and otherwise retaining the same language in the vignette might alter the findings observed in Study 1. Though we did not have a strong prediction about how Study 2 might differ, we considered the possibility that simply *naming* the slurs—particularly for misogyny and hybrid—might cause them to be rated as "worse" across conditions.

Participants. 200 participants recruited from Amazon's Cloud Research completed the experiment in a median time of 6.4 minutes and were compensated $1.50 for their time. No participants were excluded from analyses. The sample had a mean age of M=37.9 (SD=10.6); 106 were female.

Design & Procedure:

Once again, Study 2 used a 1(Agent/James: white man who loses his temper) x 4(Target/Taylor: White woman—misogyny condition vs Black man—racist condition vs(Black woman—racist + misogyny condition vs White man—control condition) full-factorial, between-subjects design. In everything but some slight modification of language in the vignette (and the change in the "misogyny" question noted above), the two studies were identical. Participants were again randomly assigned to one of the four conditions, which each contained approximately 50/cell. Participants then read a brief vignette in which a "target," always named Taylor was openly and frequently disparaging of the "Agent," who was always named "James." The scenario then describes how James reacts by losing his temper and in the course of it calling "Taylor" pejorative terms. Importantly, in Study 2, the three conditions stipulated that 1)in the misogyny condition, the Agent used "sexist slurs," 2) in the racism condition, the Agent used "racist slurs," and 3) in the hybrid condition, the Agent used "sexist and racist slurs". The control condition was worded exactly as it was in the previous experiment. The precise wording for all conditions is presented below in Table 1.3:

TABLE 2.4. Description of the manipulation in Study 2. James, the offender who uses vicious language is always a White man and is bolded, as is the control version of Taylor; Taylor, the target, is a different demographic in each condition and is color-coded. These font deviations were not used in the presentation to participants.

All conditions first began with the following phrasing. The two misogyny conditions used female pronouns, of course, while the other two conditions used male pronouns when referring to the target.

You work in the same department as two people, **James** and Taylor. Taylor is known to deliver tough, blunt criticism and by general reputation does not suffer fools. Most people know that he (or she) doesn't particularly respect **James**, whom he (or she) regards as a poor

From here, the vignettes diverged according to condition:

TABLE 2.4 (continued):

MISOGYNY CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses **his** ideas as unproductive and stupid. In a moment of aggravation, **James**, **a White man,** loses his temper and viciously insults **Taylor, a white woman,** using **sexist slurs.** **He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

RACIST CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses **his** ideas as unproductive and stupid. In a moment of aggravation, **James, a White man,** loses his temper and viciously insults **Taylor, a Black man,** using **racist slurs**. **He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

MISOGYNY & RACIST CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses his ideas as unproductive and stupid. In a moment of aggravation, **James, a White man,** loses his temper and viciously insults **Taylor, a Black woman,** using **sexist and racist slurs.** **He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

CONTROL CONDITION:

One day during a project meeting, **Taylor** targets **James** especially intensely and dismisses his ideas as unproductive and stupid. In a moment of aggravation, **James,** a **White man**, loses his temper and viciously insults **Taylor,** who is **also a white man,** calling him names in the course of it. **He** later apologizes and says that **he** really lost it after having undergone such negative treatment and that it won't happen again.

Study 2 Results

We found substantial support for all three hypotheses, yet not to the same degree as in Study 1. Explicitly identifying the slurs as racist or sexist be had different effects upon each condition—and was not consistent across DVs. Once again, we used the Holm- Bonferroni

method for multiple comparison. Further details about the general results for each DV are available in Supplemental Analyses.

Hypothesis 1; "Bigotry Light"*:* Racism is more offensive and meaningful than misogyny.

Hypothesis 2; "Two Wrongs"*:* Rather than compounding judgments of racism, the hybrid condition (racism + misogyny) will be rated as less offensive than racism, alone, and will likely fall somewhere between the two other experimental conditions (pure racism and pure misogyny).

Dependent Variable: Bigotry

As depicted in Figure 2.4A, a one-way ANOVA yielded a main effect of bigotry, $F(3, 194)=16.0$, $p<0.001$, $\eta^2=0.198$ and an adjusted coefficient of determination, $R^2=0.156$. While the Agent in the racism condition ($M=70.6$, $SD=25.3$, 95%CI=[63.4-77.8]) was deemed *marginally* more of a bigot than the Agent in the misogyny condition: ($M=62.6$, $SD=24.2$, 95%CI=[56.0-68.2]; $t(96.5)=0.167$, $p=0.098$, $d=0.335$), this finding did not, in itself, provide strong evidence for the Bigotry-Light Hypothesis. In marked deviation from the previous study, the hybrid condition ($M=74.7$, $SD=19.8$, 95%CI=[67.5-81.3]) was rated highest. Crucially, however, the hybrid condition did not differ from the racism condition, ($t(88.9)=0.836$, $p=0.405$, $d=0.171$) while it *did* differ from the misogyny condition, ($M=62.6$, $SD=24.2$, 95%CI=[56.0-68.2]; $t(96.7)=2.76$, $p=0.007$, $d= 0.548$). The fact that the Agent in the hybrid condition was rated substantially more of a bigot than the Agent in the misogyny condition but not the racism condition lends indirect support to the Bigotry-Light Hypothesis. Further, while the hybrid condition was not *lower* than racism, it did not differ—and was far from additive—and, therefore, coincided with Two-Wrong's Hypothesis. All conditions differed from control,

M=44.6, *SD*=24.1, 95%CI=[37.9-51.3]) including misogyny. The pairwise comparison between those misogyny and control was the following: $t(99.9)=3.75$, $p<0.001$, $d=0.742$.

A) B)

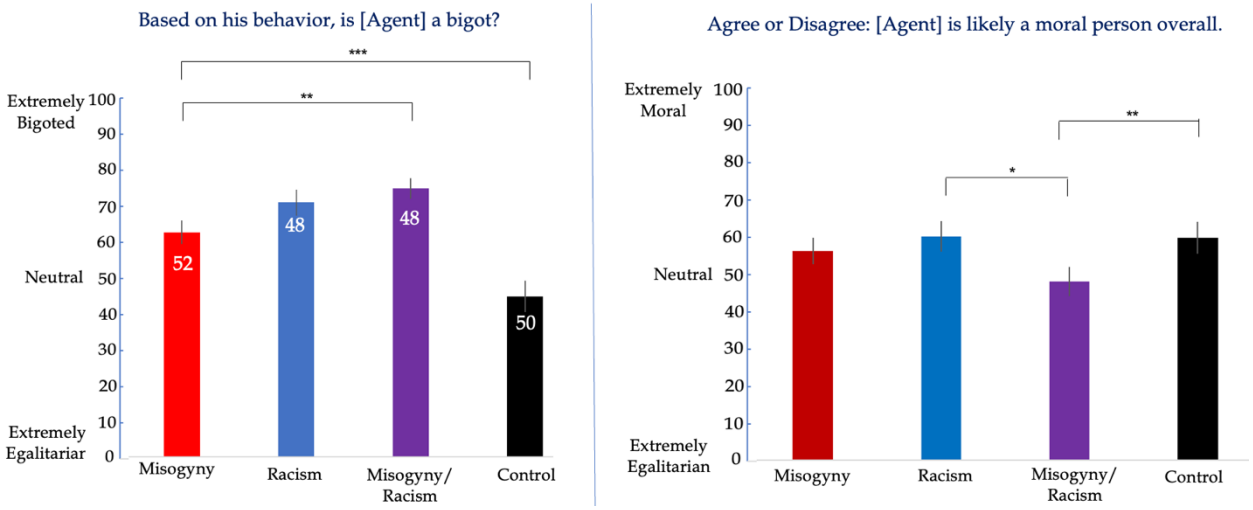

Figure 2.4. Evaluations of A) characteristic bigotry of the Agent; B) and characteristic morality of the Agent. Titles are the respective questions presented to the participants, and bars represent SEs. As represented in Figure 2.4A, $n=52$ for the misogyny condition, $n=48$ for the racist condition, $n=48$ for the hybrid condition, and $n=50$ for the control condition.

Dependent Variable: Morality

As depicted in Figure 2.4B, a one-way ANOVA yielded an effect of morality, $F(3, 197)=2.97$, $p=0.033$, $\eta^2=0.044$ and an adjusted coefficient of determination, $R^2=0.023$. In contrast to the moral findings in Study 1, the model was barely significant by Holm Bonferroni correction, and the pattern was very different. The Agent in the hybrid condition, ($M=47.3$, $SD=26.3$, 95%CI=[39.8-54.8]), was rated least moral and significantly differed from the racism condition, (M=59.4, SD=28.0, 95%CI=[51.5-67.3]; $t(93.7)=2.17$, $p=0.032$, $d=0.443$), as well as control, ($M=61.3$, $SD=19.6$, 95%CI=[55.9-66.7]; $t(86.7)=2.95$, $p=0.004$, $d=0.600$). No other pairwise comparisons were significant, including that between misogyny and racism, which were almost even.

In this case, the hybrid condition was greater than the racist condition, though not the misogyny condition, which still arguably conforms to the Two-Wrongs Hypothesis. As discussed in detail earlier, one might expect that in the combination of racism and misogyny, the hybrid condition would be additive. Certainly, it logically should be different from *both* the rating for racism and misogyny —and the effect sizes should be large. We, of course, predicted that none of those points would be true, largely on the basis that people consider racism to be worse than misogyny. If our interpretations are correct, Study 1 conformed to all three of our hypotheses. However, it's intriguing that what appear to be small differences in semantics have clearly exerted an impact on participant construal and evaluations of the respective Agents in the scenarios. Even more notably, the effects on the different conditions do not appear to be the same—or even entirely consistent across DVs. As tentatively predicted and will be discussed further, misogyny and hybrid were both—for most DVs—rated to be "worse" than their counterparts in Study 1. However , for this particular DV, *the only condition that changed appreciably* was the racist condition, which was much higher than its counterpart in Study 1 at M=37.3; the other three are virtually the same. The difference between this condition in Study 1 (M=37.3) and Study 2 (M=59.4) came to: $t(89.1)=4.21$, $p<0.001$, $d=0.867$. Somewhat expectedly, the framing of the "slurs" used seemed to have had the opposite effect on the racism condition, apparently reducing the impact of the Agent's experience on participant appraisals. For almost every DV, racism was rated "better"—in magnitude— sometimes significantly and substantially than its counterpart in Study 1, especially in comparison to the hybrid condition. It also, of course, served to change the pattern of results. However, we reserve further discussion for a larger review of the dependent measures.

Dependent Variable: Goodness of Action

The main effect of goodness-of-action was not significant, $F(3, 194)=2.14$, $p=0.097$, $\eta^2=0.03$, with an adjusted $R^2=0.017$. Though there were no significant comparisons, the hybrid condition, ($M=31.2$, $SD=22.2$; 95%CI=[14.0-26.6]) was marginally different from the three other conditions: racism, ($M=47.5$, SD=34.5,95%CI=[37.9-57.2]), misogyny, ($M=42.9$, $SD=33.5$ 95%CI=[33.8-52.0]), and control, ($M=42.8$, $SD=27.9$, 95%CI=[36.1-49.3]).

We include this graph and expound on this DV largely as a particular contrast to the findings in Study1. In this instance, the Agents in all experimental conditions were rated higher—their behavior considered better—than their counterparts in Study 1. This phenomenon was especially true of the racism condition, which went from a mean of 13.3 in Study 1 to one of 47.5 in Study 2. Once again, in contrast to our predictions, using what we regarded as more explicit language had the opposite effect from what we loosely anticipated, which is that people would generally rate a given Agent's behavior and character as either the same or a bit more harshly for most DVs. Instead, it appears that "racist slurs" (Study 2) rather than "slurs applied only to Black people"(Study 1) may have seemed, ironically, more diffuse and less serious. Something in the language of "slurs only applied to Black people" might trigger strong reactions for two reasons: it emphasizes again that it's *black* racism and also suggests there's something unspeakable about it, prompting thoughts, for instance of the—N-word—the only slur that people who consider themselves remotely egalitarian—or want to appear so— will not dare say, even in intellectual discourse. Once again, there's no such word for misogyny. We expound upon these points in later discussion.

A) B)

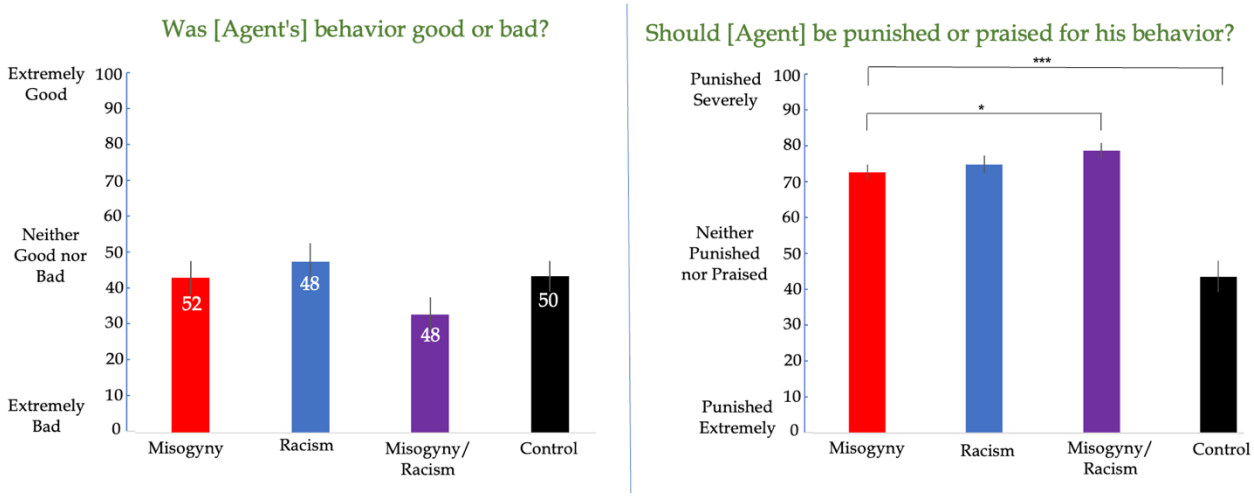

Figure 2.5. Evaluations of A) the good or bad behavior of the Agent; and B) ratings of whether the Agent should be punished. Titles are the respective questions presented to the participants, and bars represent SEs. As represented in Figure 2.5A, n=52 for the misogyny condition, n=48 for the racist condition, n=48 for the hybrid condition, and n=50 for the control condition.

Dependent Variable: Praise/Punishment

A one-way ANOVA yielded a main effect of "punishment," $F(3, 194)=10.1$, $p<0.001$, $\eta^2=0.135$, with an adjusted $R^2=0.121$. The hybrid condition, ($M=78.2$, $SD=15.1$,95%CI=73.9-82.5]) was highest in magnitude and just short of significance in comparison to misogyny, ($M=72.1$, $SD=15.2$, 95%CI=[68.0-76.2]; $t(97.4)=1.96$, $p=0.052$, $d=0.393$), though not racism, ($M=74.4$, $SD=17.1$, 95%CI=[69.6-79.2]; $t(92.7)=1.13$, $p=0.260$, $d=0.231$). All of the experimental conditions differed from control, including misogyny: $t(75.1)=7.08$, $p<0.001$, $d=1.42$.

Contrary to many of the other DVs, in this case, racism did not differ much from its Study 1 counterpart. As is apparent in Figure 2.5B**,** the hybrid condition was greatest in magnitude—a 78 in Study 2 where it had been a 70.1 in Study 2. To be clear, in both studies, the experimental conditions were near ceiling, so this measure appears to register that participants believed use of slurs warranted punitive action to a much greater degree than for the use of

"uncharged" insults (control). There is still evidence for "Bigotry-Light" but the effect—since the Hybrid differs from misogyny but not racism—but it is more subdued. There was also arguably support for the "Two-Wrongs," as in most cases the hybrid condition did not significantly differ from the racist conditions—with a few notable exceptions.

Hypothesis 3; "The Gateway:" Once again, theorizing that sexism is generally regarded as less offensive and threatening, we predicted that people might infer misogynistic tendencies even when provided no such information and only evidence of racism. Notably, we expected the inverse to be less common: we did not expect people to infer racism from misogynistic information. Study 1 seemed to provide support for both aspects of this hypothesis. Intriguingly, Study 2 added nuance. The change in semantics suggested that, in fact, people may—under certain circumstances—make the opposite assumptions as well: that misogynistic propensities can likewise imply that one has a greater likelihood of racist characteristics. It appears that the first inference—racism implying misogyny is greater and probably more common—but we discuss in further detail below.

Dependent Variable: Sexism—

A one-way ANOVA produced a main effect of sexism, $F(3, 194)=17.2$, $p<0.001$, $\eta^2=0.210$, with an adjusted $R^2=0.198$. Interestingly, all three experimental conditions—and, in fact, control, which was worded exactly the same in both studies—were much higher in magnitude, particularly misogyny, (M=73.6,SD=21.0, 95%CI=[67.9-79.3]) and hybrid, (M=73.4 SD=20.6, 95%CI=[67.6-79.2]), which did not differ. Both differed from racism, (M=59.0 SD=27.6, 95%CI=[51.2-66.8]). Pairwise comparison with misogyny produced, $t(87.7)=2.96$, $p=0.004$, $d=0.599$; comparison with hybrid yielded, $t(87.0)=2.68$, $p=0.009$, $d=0.547$. The Agent

in the racist condition was rated significantly more likely to be sexist than the Agent in the control condition: M=43.3, SD=24.8,95%CI=[36.4-50.1]; t(94.0)=2.84, p=0.006, d=1.29.

Similar to bigotry—in which hybrid and misogyny were rated much higher than in Study 1, the sexism DV was another example in which these two conditions changed more dramatically than racism. However, both racism and control were also rated higher than their Study 1 counterparts. Intriguingly, this is closer to the pattern of results we initially expected for this DV in Study 1, as in that case, misogyny and hybrid were barely above the midline and did not even differ from racism. This time, the "sexist conditions"—misogyny and hybrid—were much higher than racism, as exhibited by robust effect sizes. Their means fall within the 70's, which is much closer to where people place the "racism"—though not hybrid—in Study 1 in the racism DV. However, as in Study 1, however, there was, once again, a strong difference between the racism condition and the control, even though the racism scenario made no mention of any sexist tendencies—or any such mild reference at all. Further, racism differed significantly from the midline into misogynistic: t(49)=2.27, p<0.028. Once again, this supports The Gateway Hypothesis that people infer misogynistic characteristics from racist information.

A) B)

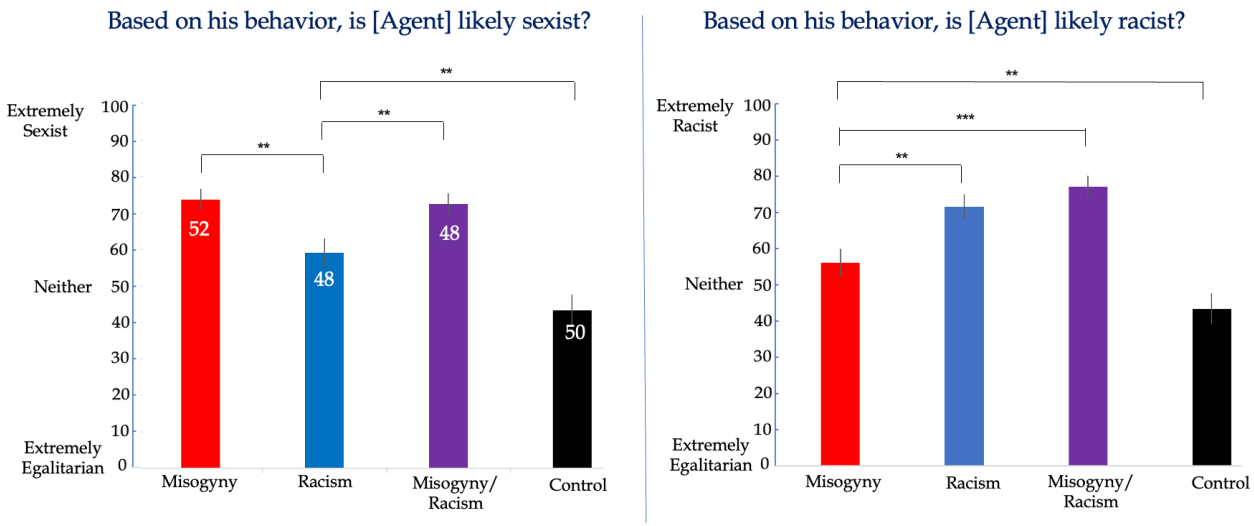

Figure 2.6. Evaluations of A) the sexist characteristics of the Agent; and B) racist characteristics of the Agent. Titles are the respective questions presented to the participants, and bars represent SEs. As represented in Figure 2.6A, n=52 for the misogyny condition, n=48 for the racist condition, n=48 for the hybrid condition, and n=50 for the control condition.

Dependent Variable: Racism

A one-way ANOVA produced a main effect of racism, $F(3, 194)=17.0$, $p<0.001$, $\eta^2=0.243$, with an adjusted $R^2=0.231$. Once again, the racism condition did not differ much from Study 1, while hybrid was much higher in magnitude in Study 2 than in Study 1, as exhibited by Figure 2.6B. However, there was, yet again, no difference between hybrid, (M=77.0, SD=21.2, 95%CI=[71.02-83.0]), and racism, (M=71.5, SD=23.5, 95%CI=[64.9-78.1]; $t(93.0)=1.20$, p=0.233, d=0.245. While hybrid was higher in magnitude, this is further evidence for the Two-Wrongs Hypothesis. As expected, both hybrid and racism were much greater than misogyny, (M=56.1, SD=27.6, 95%CI=[48.6-63.6]). Comparison between hybrid and misogyny produced: $t(94.9)=4.27$, $p<0.001$, d=0.847; comparison between racism and misogyny yielded: $t(97.3)=3.02$, p=0.003, d=0.602. Unlike in Study 1, the Agent in the misogyny condition was rated to be more likely to embody racist elements than the control condition: M=40,

SD=22.1, 95%CI=[22.0-34.4]; *t*(99.8)=2.82, *p*=0.006, *d*=0.559. However, unlike the racist condition in the "sexist" DV, misogyny did not differ from the midline: *t*(51)=1.58, *p*=0.120. Further, the effect size between racism and control was much greater than that between misogyny and control.

Dependent variables: Jerk, Friend, Aggression, Respect

Once again, we did not include graphs for these measures, but they exhibited many of the findings that permeated the DVs in Study 2. Clearly, the results were far less consistent that in Study 1, and they also did not coincide with our tentative predictions that 1) the differences in language would not produce large differences from Study 1, and 2) that in all cases, they would not necessarily cause people to rate the experimental conditions more forcefully. The case of the "jerk" DV as well as the "respect" DV—as in the "bigotry" DV— were two examples in which the racism condition did not seem to change much in magnitude while the hybrid and misogyny conditions became substantially greater in magnitude, as compared to Study 1. The "respect" DV is, perhaps, especially revealing. As a reminder, the higher the rating, the more the participant regarded the given Agent as failing to respect "those who are different from him." "Bigotry" and "respect" DVs suggest that simply naming the slurs, "sexism" and "sexism and racism," respectively, endowed the respective treatments with much more force.

By contrast, with regard to the "friendship" DV, the other conditions remained relatively similar to their Study 1 counterparts, while the Agent in the racism condition evidently became more eligible as a possible friend. Even more startling, another example in which the racism condition alone seemed to dictate results was "goodness-of-action" and "morality": As discussed in the context of the "morality" DV, the Agent in the racism condition was rated *M*=37.3 in

Study 1 and *M*=59.4 in Study 2. "Aggression" for all conditions was very high, so there was not much to extract in examining the differences between the studies.

Study 2 Discussion

Before considering the marked differences between Study1 and Study 2, Study 2 largely supported our hypotheses, though with added nuance. Of the three, Study 2 probably provided the most support for the Two-Wrongs Hypothesis and indirectly, the Bigotry-Light, as the Hybrid condition tended to differ from misogyny but not racism. While racism never differed from misogyny, the hybrid condition frequently did and, critically, did not differ from racism. This combination provides support for Bigotry-Light, though tempered by what we were clear differences in how people were interpreting the scenarios. Intriguingly, it seems that while misogyny may often be more tolerated and may lead to other forms of bigotry, people may also assume that if you're strongly misogynistic, you are more likely to be racist.

General Discussion

For Study 1, the results were remarkably consistent across DVs, and it would be difficult to conjure a competing interpretation that would account for results, where 1) Racism was *always* worse than misogyny with large effects, and 2) the combination in the form of a black woman never exceeded racism. At the very least, one can conclude that, quite literally, "viciously" insulting a black man with a barrage of "slurs applied only to Black people" is far worse than "viciously" insulting a white woman with a barrage of "slurs applied only to women." Study 1 provides strong support for all three hypotheses. Study 2 provides nuance. We hazard some theories, but only further investigation will illuminate the cognition underlying

what initially appeared a rather slight tweaking of language between the two studies and somehow yielded some large differences.

The vagaries that emerged in Study 2 are not necessarily obvious or consistent, as noted. However, in combination, these findings largely suggest that 1) the hybrid and misogyny conditions were considered more meaningful and serious in Study 2, and 2) the racism condition largely experienced the opposite effect and was not considered as serious or repugnant in comparison to Study 1. While further research is required to explore the mechanisms behind these evaluations, several possibilities immediately present themselves. With regard to the racism condition, it may have seemed that explicitly naming the slurs as "racist" would impress people more forcefully with the seriousness of the bigotry. Yet for the racism condition, it may (and seems to) have had the opposite effect. For the precise reason that is central to our main hypothesis—that black racism is considered far worse than misogyny—even using the phrase "slurs only *applied to Black people"* may have evoked the darkest forms of black racism: The N-word, the violence, the terrible history of violence, the continued discrimination. Without realizing it—and, of course, in hindsight's conjecture—we may perhaps have *more* effectively elicited the essence of racism, rather than simply stating that "James" used "racist slurs," In other words, avoiding the use of any term and simply stating, "only applied to black people" may have called to mind all that some regard as unspeakably horrific about black racism, past and present. To state the obvious, there's no doubt that there was a tempering of how people were interpreting the racism the target experienced, so it must have occurred through this difference in language.

In stark contrast, stating that a white woman experiences slurs applied only to women in Study 1 probably—more predictably—had largely the opposite effect. Again, the reasoning relates to our central hypotheses that people struggle with misogyny in the abstract: 1) People

don't generally believe women can really be a psychological minority; 2) slurs—along with most other aspects of being a minority—aren't very serious with regard to women. For a very public representation of this likelihood, the comedian Samantha Bee called Ivanka Trump a "cunt"—she did later apologize—yet in the meantime was applauded by all of Hollywood and much of the "left" who regard themselves as egalitarian; and perhaps most importantly, 3) People have particular difficulty imagining that *white* women can experience true bigotry—again, a reiteration of our central hypothesis. In short, people fail to comprehend—or brave the act of admitting— the breadth of misogyny through the thick veil of "whiteness" of which they are also possessed—and from which they, of course, benefit. As noted, they gain from being white, just as black men gain from being men, as discussed in detail. In neither case is the bigotry diminished. Leaving people with an open idea of "slurs applied on to women" probably conjured mixed feelings of what that even entails. As earlier discussed, when people presume to talk about misogyny in isolation, that discussion is often swiftly overwhelmed by intersectional points, for one. Indeed, even daring to discuss misogyny often ironically invokes use of that misogyny. In fact, one participant in this experiment who was assigned to the sexist condition referred to the—in this case white, female—victim of the vignette as a "bitch" in the comments sections: "Yeah, but Taylor was also a bitch." In addition, though this is retrospective hindsight, we suspect that naming "sexist slurs"—especially in an office environment—probably provoked the one kind of bigotry that people have started to recognize: sexual harassment and related violence. In the case of the hybrid condition, naming both "sexist" and "racist" slurs seems even more so to have provided a clear sense of the target's treatment, which in the first study was much more vague.

In broad conclusion, we set out to explore misogyny by comparing it to black racism. Study 1 provided strong evidence for our hypotheses that racism is considered more serious and

meaningful than misogyny, that black women occupy a strange limbo where the full impact of their plight does not seem to register, and that misogyny may lead to other forms of bigotry. However, Study 2 provided a powerful reminder of how complicated bigotry can be if even apparently simple differences in semantics yield different results. Once again, we expect that racism is considered much worse than misogyny but partly because people don't believe it occurs as often or has as wide a breadth as, say, black racism. These questions would benefit from further research into what how people think about the *content* of sexism, racism, and other forms of bigotry.

Having compared misogyny and racism, we proposed to explore misogyny within and through the context of one of the single closest relationships: the heterosexual marriage of a man and woman. While the evaluations relate to these two people and how each one's decisions affect the other, society inevitably intrudes. Chapter 3 provides some revealing insights what one of the closest relationships on Earth can reveal about power and "otherness."

CHAPTER 3:
AS LONG AS MEN ARE BETTER THAN WOMEN—EXPLORING GENDER NORM-VIOLATION WITHIN HOME AND WORK LIFE AND HOW THEY ARE AFFECTED BY PRESTIGE, SALARY, CAREGIVING, AND PARTNER POSITIONS

INTRODUCTION

One of the artifices of Satan is, to induce men to believe that he does not exist: another, perhaps equally fatal, is to make them fancy that he is obliged to stand quietly by, and not to meddle with them, if they get into true silence.

—Charles Baudelaire, “Le Figaro,” 1864

Misogyny is the devil in the canon of bigotry—particularly when anyone dares to consider it as a category unto itself. Inevitably, the dialogue turns to intersectional issues, which—while important—are different, if related, considerations. In several ways, misogyny–or the systematic prejudice and discrimination against women based on their perceived membership in this gender orientation—has remained stubbornly insidious, surviving—and thriving—into the present day. Despite being the oldest form of bigotry, it continues to battle for basic recognition. Yet, we argue that it infiltrates many aspects of daily life to the great detriment of both men and women (Momennejad et al., 2019). The literature supporting the gravity and extent of misogyny's reach is mounting.

In particular, there has been increasing exploration of norm violation both inside and outside the traditional home and how this affects the existence and perpetuation of discrimination against women (e.g. Gaunt, 2013). Until recently, research investigating such questions in a naturalistic environment had been relatively rare. Discrimination and prejudice arguably begin in the unit of the household, where men—particularly those who feel threatened by women (e.g., Cross et al., 2019)— still maintain psychological (and at times, physical) dominance that extends into the public sphere. To be clear, men don’t have to impose literal physical violence to assert their dominance. Even brandishing their physical strength can have a detrimental effect on a partner that affects how she interacts with the rest of the world—especially men—professionally and socially. However, it’s not just insecure husbands who perpetuate this discrimination, of

course. More importantly, there are powerful societal norms at the intersection of work and home that men and women internalize early in life—in a word, the patriarchy. This patriarchy remains strong for many reasons, but not least of which is that women are perpetually second class. Unlike all other minorities, there is almost never a context in which women have systematic ascendance.

Women—and men—are taught early about relative female incompetence, and especially relevant to this study, the subtle, insinuating fact that they are not quite people of the world. For instance, consider the "work" of cooking vs the "craft" of cooking—or, in other terms, the home cook vs. the public chef. Historically and to a large degree in the present day, household cooking has been the purview of women: it's associated with duty and "Grandma's secret recipes" but is most certainly *not* associated with craft, expertise, and most importantly, public prestige. The latter are the characteristics of the chef, who is decidedly male. The material point is that once something moves from the household into the public or technical sphere, it apparently ceases to be "feminine" and becomes "masculine." Indeed, scholars have long reported that women are favored when they conform to stereotypes and suffer condemnation when they are perceived to defy them (Glick & Fiske, 2001). For instance, some have argued that sex differences are largely dictated by staunchly ingrained beliefs regarding gender roles (e.g., Eagly & Wood, 2000).

However, while social roles undoubtedly contribute to bigotry against women—and vice versa—there is, nonetheless, discrimination underlying and interacting with this relationship. In our current research, we seek to expand on social roles and normative theories and also demonstrate that other factors are at work. As one of the most revealing examples of how discrimination extends beyond differing roles, children as young as six-years-old believe—like their adult counterparts—that "being really smart" or "being brilliant" is the mandate of boys

(e.g., Bian et al., 2017). Such beliefs have cascading effects as stereotypes about female's relative incompetence range across a broad swath of skills and endeavors. This discrimination has life-long effects on women, their career choices (e.g., Rudman & Glick, 1999), their engagement in activities deemed to be masculine, and on their basic self-esteem and self-efficacy (e.g., Dickerson & Taylor, 2000).

The Current Research

There is a particular resistance against acknowledging that women are a psychological minority and experience grave discrimination and prejudice. Consider that there are extremely few—if any—circumstances in which women have control. She is perpetually second class, from household to the broader world. The patriarchy—dominant in every context—has little incentive to cede any power. In fact, as research suggests, many men fear the ascendence of women (Cross et al., 2019). Yet, as is obvious, male and female lives are deeply intertwined—unlike no other two groups. The current study thus proposes to explore how control, power—and in particular *choice*—affect participant evaluations of decisions that men and women make for the sustenance of work and home. We proposed to combine male and female decisions into the same conditions so as to further consider how partner situations and decisions add contextual nuance to these appraisals. We know of no other studies that have factored in the interactional context of a husband and wife making decisions that we expect will affect construal of the same actions if there were no partner under consideration.

In this study, we thus sought to explore what the same decision-making–within the shifting context of work and home life–reveals about the stalwart nature of gender roles and stereotypes and discrimination. In a 2(gender: man vs. woman) x 3(condition: voluntary

surgeon/voluntary nurse vs. involuntary surgeon/stay-at-home parent x voluntary nurse/voluntary stay-at-home parent) design, we, once again, seek to examine how people differentially evaluate gender decisions, particularly when intellectual prowess, *prestige*, and, *choice,* are manipulated. The conditions incorporate how people perceive gender in the context of prestige and public responsibility (male) and private homelife and child-care (female). Critically, we also explore how *choice* affects these dynamics. We predict that men forgive and even endorse women who are deprived of choice and *compelled* to take on stereotypically masculine roles. Finally, we consider how the partner's decisions affect ratings. It's worth mentioning that a range of studies—some contradictory—argue that people apply unconscious base-rates (based on stereotypes) to men vs women so that, for instance, a female financier who makes the same amount of money as a male financier will, nonetheless, be perceived as *more* successful than her male counterpart because the expectation is lower (e.g., Biernat & Nelson, 1994). For a relevant example, one might argue that the voluntary female surgeon would be judged more highly than a voluntary male surgeon partially because of this base-rate fallacy, yet it would not explain why an *involuntary* surgeon—who has been pressured to accept this career—would be rated more highly than a voluntary surgeon, which was one of our predictions and findings. Thus while, base-rate fallacy may be a true effect—under some circumstances— it does not interfere with our primary hypotheses and predictions, which are based on exploring the same decisions under differing contexts.

Hypotheses & Design

Overarchingly, we hypothesize that women will be praised, considered a better decision-maker, regarded as more intelligent, etc. when she conforms to stereotypes regarding staying at

home to care for children or having a career that's both relatively (with respect to her husband) less lucrative and less prestigious. Conversely, we expect that men will be denigrated for a having a "lesser" public role than his wife or staying at home. There is so much to support the idea that women are associated and lauded for taking on the role of caregiver, that it hardly warrants discussion; one study recently showed that a woman was condemned for returning to work after having a baby (Coleman & Franiuk, 2011); likewise, studies support to this day that people have less approval and respect for men who stay-at-home to care for children (Coleman & Franiuk, 2011) As discussed however, we are interested in how evaluations change when circumstances and partner choices vary while protagonist decisions remain the same. Furthermore, while people may have unconscious double standards, we also argue and test the possibility that motivated cognition is involved. Men (as a group) want to preserve their superior status so have an incentive to endorse normative behavior and condemn normative-defying behavior. While women may not share this motivation, they have, nonetheless, likely internalized the effects of it.

The study is composed of six conditions. In each case, a vignette describes a husband and wife, Nick and Emily, who are both described as having a strong, early interest in medicine. In the "voluntary surgeon" conditions, Emily or Nick chooses to be a nurse in order to devote more time to the couple's child while the spouse decides to be a surgeon, despite the high demands on his or her time. This condition, therefore, comprehends more than just one person's voluntary decision to be a surgeon, it also includes the spouse's voluntary decision to work as a nurse. Crucially, the vignette indicates that the nurse's decision—whether Nick or Emily, depending on which of the two vignettes—was simultaneously motivated by a desire to have more time with their child. Further, the "ambition" of the surgeon is emphasized as we predict that evidence of

prestige and pursuit or the desire for them will be regarded less admirably in a woman than in a man. By contrast, in what we term the "involuntary-surgeon conditions," either Emily or Nick "opt" to stay home to spend more time their child. Importantly, the vignette stipulates that this decision "[pressures]" the spouse into becoming a surgeon so that there is enough money to comfortably provide for the family. In the context of this scenario, the vignette specifically states that the surgeon—e.g., Emily or Nick based on condition— has no inclination for this kind of work and accepts the position reluctantly.

Finally, there are two conditions in which Emily or Nick is either a nurse anesthesiologist or a stay-at-home parent. Collectively, the conditions were established to determine how participants would evaluate these choices—norm complying and norm disregarding—differentially for men and women, and, in particular, how choice and prestige would affect these appraisals. The two nurse vs. stay-at-home conditions were added in order to examine more closely the stigmas regarding men vs. women choosing a career once reserved for women and men staying home. More broadly, we were equally interested in the inverse of these considerations: what the choices of Emily and Nick—whom we term the protagonists—in these differing contexts reveal about status and purview in relation to gender. We hypothesize that the public sphere—complete with status, large salaries, and responsibility remain the jurisdiction of men with the home, children, and lesser jobs the purview of women.

TABLE 3.1: *Hypotheses*

Overarching Hypothesis*:*

Women will be regarded more favorably when they opt to stay home or take a "lesser" job—surgeon being "highest" followed by nurse anesthesiologist, and finally, stay-at-home as last. We expected they will be judged less positively when they take on lucrative and prestigious

TABLE 3.1 (continued):

roles, especially when they are "higher" than that of their male counterparts, respectively; the opposite will apply to men. Further, we anticipated that *choice* would feature in these decisions. For instance, we predicted that people would rate the female involuntary surgeon more highly than the female voluntary surgeon. We outline our major predictions below:

- **Prediction 1**: As just noted, the woman who takes on the role of surgeon (involuntary surgeon condition) after her husband pressures her by opting to stay at home will be rated highest among the surgeon conditions because she is compelled and, somewhat perniciously, it was not a real choice, which would be less threatening to a man. We argue she is rewarded for taking on the norm-violating role, precisely because, though she is less suited for it, she does it out of some necessity. Conversely, the woman who opts to be a surgeon will be rated lowest among the surgeons for narcissistically choosing career over time with her child, violating gender norms.

- **Prediction 2**: The man who opted to stay at home would be least praiseworthy, both for choosing not to work and for putting the onus of providing for the family on his wife; secondarily, we suspect that people would judge him for "depriving" his wife time with their child and further violating gender norms. We also speculate that he will be regarded as of inferior intellect and not contributing much to the family. Crucially, we theorize he will be judged more harshly than the other man who makes the exact same decision to stay at home but is paired with a wife who *chooses* to work as a nurse, especially since a nurse is a stereotypically "female" profession.

- **Prediction 3**: The nurse who voluntarily takes this job and 1) cites time with her child and 2) is paired with an ambitious surgeon husband will be rated highly. We also suspect that she will be rated more highly than the other nurse, who has a

TABLE 3.1 (continued):

husband that doesn't work, violating gender dynamics that the husband should have more prestige than the wife.

- **Prediction 4:** For similar reasoning outlined for the other DVs, we predict that the female surgeon who "involuntarily" violates norms and takes on roles "beyond her," the greater her perceived contribution to the marriage. We suspect that people will, once again, reward the same behavior within the context that we've already examined. For instance, the female involuntary surgeon—working hard against her inclination and against her nature as a woman—would be regarded as contributing especially highly to the marriage.

- (**Speculative) Prediction 5**: The measure about the family money will be skewed towards women. When they don't work, they will be rated as close to sharing the money equally with men. When the do work for a higher salary than their husband, people will indicate that the money belongs more to them than their male counterparts.

OPEN SCIENCE

This study was pre-registered on The Open Science Network (OSF), which is maintained by The Center for Open Science (COS), a non-profit organization dedicated to the integrity and transparency of scientific methods and findings. We recruited participants using Cloud Research, an online platform created and supported by Amazon. We limited eligibility to people ages 18 years and over, living in in the United States, and holding a high standing within Cloud Research—meaning past work indicated that they were attentive and completed tasks in a timely

manner. Otherwise, we did not target any group or demographic or ultimately make any exclusions from the collected data.

STUDY

Methods

Participants

346 participants were recruited from Amazon Mechanical Turk using Cloud Research. The intended number was an even 300, but we ended up with an extra 46. Despite our best effort to randomize and counterbalance among conditions, occasional drop-out produced slight discrepancies in the number per cell. Participants completed the experiment in a median time of 5.35 minutes and were compensated $1.50. The sample had a mean age of M=41.7, SD=11.4 and a minimum requirement of 18 years. 145 were female. None were excluded from analyses.

Design & Procedure: Once again, each condition had two essential components, one that pertained to the woman protagonist (Emily) and one that pertained to the man (Nick). Table 3.2 provides the exact wording of each of these conditions.

TABLE 3.2: *Between-subjects conditions as the appeared to participants*:

Female, Voluntary Surgeon/ Male, Voluntary Nurse	Male, Voluntary Surgeon/ Female, Voluntary Nurse
Imagine the following scenario: Emily and Nick met at a prestigious college where they shared a passionate interest in pursuing medicine. They married early and had a child. Eventually, Nick opted to be a nurse anesthesiologist, citing more time with his child, yet still allowing for a very good salary. Extremely ambitious, Emily instead decided to pursue the lucrative and highly demanding route of being a surgeon.	***Imagine the following scenario:*** Emily and Nick met at a prestigious college where they shared a passionate interest in pursuing medicine. They married early and had a child. Eventually, Emily opted to be a nurse anesthesiologist, citing more time with her child, yet still allowing for a very good salary. Extremely ambitious, Nick instead decided to pursue the lucrative and highly demanding route of being a surgeon.

Female, Involuntary Surgeon/ Male Voluntary Stay-at-Home	Male Involuntary Surgeon/ Female Voluntary Stay-at-Home
Imagine the following scenario: Emily and Nick met at a prestigious college where they shared a passionate interest in pursuing medicine. They married early and had a child. Eventually, Nick opted to stay at home, citing more time with his child. Though not inclined for this kind of work, Emily felt financial pressure and decided to pursue the lucrative and highly demanding route of being a surgeon.	***Imagine the following scenario:*** Emily and Nick met at a prestigious college where they shared a passionate interest in pursuing medicine. They married early and had a child. Eventually, Emily opted to stay at home, citing more time with her child. Though not inclined for this kind of work, Nick felt financial pressure and decided to pursue the lucrative and highly demanding route of being a surgeon.

TABLE 3.2 (continued)

Female, Vountary Nurse Male, Voluntary Stay-at-Home	Male, Voluntary Nurse Female Voluntary Stay-at-Home
Imagine the following scenario: Emily and Nick met at a prestigious college where they shared a passionate interest in pursuing medicine. They married early and had a child. Emily opted to pursue her interest in nurse anesthesiology, which provides a very good salary, while Nick decided not to have a career and to devote his time to his child.	***Imagine the following scenario:*** Emily and Nick met at a prestigious college where they shared a passionate interest in pursuing medicine. They married early and had a child. Nick opted to pursue his interest in nurse anesthesiology, which provides a very good salary, while Emily decided not to have a career and to devote her time to her child.

Critically, there are really three separate contexts: 1) the voluntary surgeon who is paired with the spouse who is a voluntary nurse anesthesiologist, 2) the involuntary surgeon paired with the spouse who decides to stay home with the child, and 3) the voluntary nurse anesthesiologist whose spouse decides to stay home. These 3 contexts have male and female versions in which the roles are switched, creating a 2(gender) x 3(context) between-subjects design.

However, because some DVs have both a male and a female version, for those measures we first ran a mixed-effects regression with gender (male vs. female) as our within-subjects variable x context, creating twelve condition-responses. In other words, each participant was contributing two observations, one for the female half of their given condition and one for the male, creating twelve condition-response pairs so that we could compare every iteration.

We performed this analysis for all dependent variables that included two questions—one about Emily and one about Nick—i.e., praiseworthiness, intelligence, and decision-making. To

illustrate what might seem counterintuitive, consider the following reiteration of conditions, which consistent of six conditions but twelve contexts:

TABLE 3.3: *List of conditions and observations for each participant*

Condition 1— 2 responses: Female voluntary surgeon & male voluntary nurse =2 observations (female/male)

Condition 2—2 responses: Male voluntary surgeon & female voluntary nurse= 2 observations (male/female)

Condition 3—2 responses: Female involuntary surgeon & male insists on staying at home) =2 observations (female/male)

Condition 4—2 responses**:** Male involuntary surgeon& female insists on staying at home) =2 observations (male/female)

Condition 5—2 responses: Female nurse (male stay-at-home) =2 observations (female/male)

Condition 6—2 responses**:** Male nurse (female stay-at home) = 2 observations (male/female)

The rest of the DVs did not have a male and female component so we performed straightforward two-way ANOVAs in a 2(gender: male vs. female) x 3(context) between-subjects models**.**

Results.

Dependent Variable: Praiseworthiness/Blameworthiness

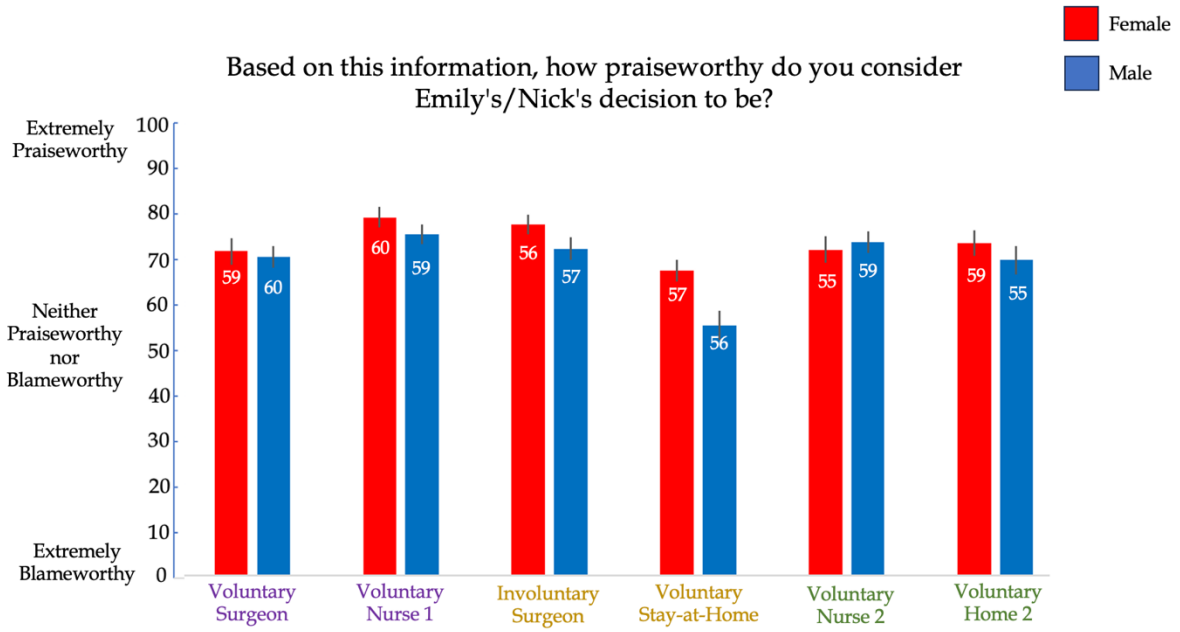

Figure 3.1. Depiction of praiseworthiness represents all six conditions and twelve contexts. Bars of the same color represent contexts that were combined into one condition. For instance, take the purple bars: the red, female voluntary surgeon was paired with the *blue,* male voluntary nurse in the condition in Table 3.1. The numbers printed on the bars refer to the n for each condition. Error bars represent *SE*.

A two-way ANOVA produced a main effect of gender, ($F(1,231)= 26.1$, $p<0.001$) and a main effect of context ($F(5, 517)=12.2$, $p<0.001$). However, the interaction between the two was not significant, though it was marginal: $F(5, 517)=2.09$, $p=0.065$).

In order to test our primary hypotheses, we ran several pairwise t-tests for ease of interpretation. With reference to the surgeon scenarios, as expected, the female involuntary surgeon was rated most praiseworthy among the four at $M=77.5$, $SD=16.1$, 95%CI=[71.2-84.0]. It differed significantly only from the male voluntary surgeon, which was rated lowest in magnitude ($M=70.5$, $SD=18.4$, 95%CI=[65.9-75.1]; $t(114)=2.19$, $p=0.031$, $d=0.405$. Interestingly, male and female voluntary surgeons did not differ from one another, suggesting that these effects are not simply a function of baseline expectation of competence. As discussed, other scholars have proposed that women are generally held to a lower standard so when they perform comparably to a man, they are, nonetheless, regarded as having exceeded him (e.g.,

Biernat et al., 1991). Instead—at least in addition to that effect—people seemed to register particular support for the woman who is compelled to take on a prestigious and lucrative role that should be reserved for a man. We argue she was not quite *choosing* this role, and this fact increased her appeal. The female volunteer and pressured surgeons did not differ, though they were just short of marginal: $t(105)=1.64$, $p=0.103$, $d=0.304$.

This interpretation regarding choices acquires further support when one considers how people rated the roles that were paired with the surgeon conditions. The female nurse whose partner was an ambitious, voluntary male surgeon, ($M=79$, $SD=17.6$, 95%CI=[74.5-83.4]) was rated highest of all twelve conditions. As a reminder, it was also mentioned that she factored time with her child into her decision to be a nurse. Comparison with the voluntary female nurse who is paired with a volunteer stay-at-home partner—one who cites a desire to spend time with his child—yielded a difference that was just shy of significance: ($M=71.9$, $SD=21.0$, 95%CI=[66.3-77.4]); $t(106)=1.94$, $p=0.055$, $d=0.365$. Though both take the same position—nurse anesthesiologist—in the first case 1) Emily is supporting what might appear a more dominant husband and 2) also expresses interest in time with her child, which seems to account for the difference.

In conjunction with other data, the fact that the nurse with the surgeon husband is rated higher (in magnitude) suggests that people favored women when they, themselves, conformed to stereotypical norms—i.e., prioritizing raising children and being submissive and unambitious. However, critically, they were lauded when their behavior facilitated the dominance and prestige of their spouse, whom norms dictate should have higher public status. These findings are further bolstered by perhaps our most robust—and consistent across DVs—results: participant evaluations of the man vs. woman's respective *decisions* to stay at home to care for their child.

Comparison of the male, (M=55.1, SD=24.2, 95%CI=[48.8-61.4]), and female caregivers (M=67.3, SD=18.5, 95%CI=[62.5-72.1]), who were paired with a surgeon were robustly significant: $t(103)=3.01$, $p=0.003$, $d=0.568$. Once again, the vignettes had implied that these caregivers had insisted on staying home, "pressuring" their spouse to take on the role of a surgeon against his or her respective inclination—or even her nature as a woman. The woman was not rated nearly so poorly, most likely because she was supporting a prestigious husband and taking on a normative role. Notably, the same decision to stay home when the caregiver was paired with a spouse who seemed content with his or her decision work as a nurse was produced a much smaller difference between the male and female counterparts. In this case, the male, (M=69.6, SD=23.4, 95%CI=[63.4-75.8]), and female, (M=73.3 SD=21.7, 95%CI=[-78.9]), stay-at-home caregivers were not significantly different, though were in the predicted direction: $t(110)=0.891$, $p=0.375$, $d=0.193$. Finally, of course, comparing the two male-stay-at-home conditions yielded a robust significant difference: $t(109)=3.21$, $p=0.002$, $d=0.609$**.**

Dependent Variable: Intelligence

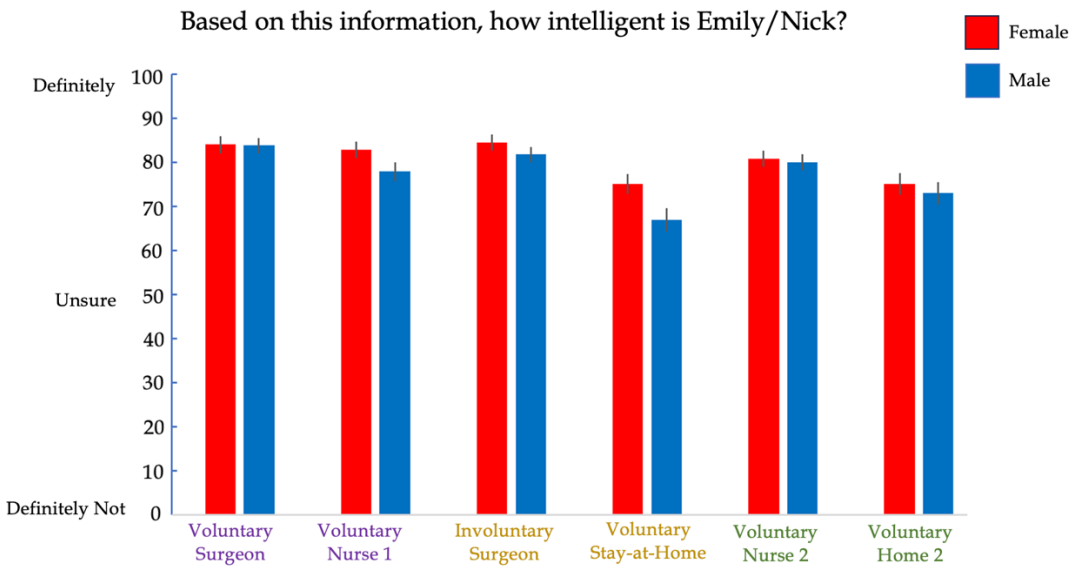

Figure 3.2. This depiction of intelligence represents all six conditions and twelve contexts. Bars of the same color represent contexts that were combined into one condition. For instance, take the purple bars: the red, female voluntary surgeon was paired with the *blue,* male voluntary nurse in the condition in Table 3.1. Errors bars represent *SE*.

A two-way ANOVA produced a main effect of gender, ($F(1, 517)=10.2$, $p<0.001$, and a main effect of context ($F(5, 517)=15.6$, $p<0.001$. However, the interaction between the two was not significant: $F(1, 517)=0.993$, $p=0.421$.

The results for intelligence produced a similar pattern to that of praiseworthiness, though with some notable differences. In this case, there was very little difference among the surgeons, which were relatively even. The predicted difference between the involuntary female, ($M=84.5$, $SD=13.4$, 95%CI=[80.9-88.0]), and male, ($M=81.7$, $SD=13.7$, 95%CI=[78.2-85.1]), surgeons was different only in direction and magnitude: $t(111)=1.10$, $p=0.275$, $d=0.206$. It's possible—given the nature of the work, surgery, that we were encountering ceiling effects. Given this possibility, it's perhaps worth noting that the female nurse, ($M=82.8$, $SD=13.7$, 95%CI=[78.2-85.1]), who supports her ambitious husband in his role as a surgeon was rated as more intelligent—though only marginally— than her male counterpart: $M=77.9$, $SD=16.8$, 95%CI=[73.6-82.1]; $t(112)=1.77$, $p=0.079$, $d=0.326$.

Notably, the man who chooses to stay at home, (M=66.9, SD=19.9, 95%CI=[61.7-72.1]), pressuring his wife to become a surgeon was considered significantly less intelligent than his female counterpart: M=75.1, SD=17.3, 95%CI=[70.6-79.6]; t(108)=2.31, p=0.022, d=0.326**.** One possible explanation for why this effect might translate to intelligence as well—which is less clear than say the reasoning for praiseworthiness—is that people can't imagine a respectable, capable man deigning to cede so much apparent power and prestige to his wife. As already noted**,** the wife may be generally faulted for evidently applying pressure on her husband, yet people still rate her as fairly intelligent for assuming a role that is perfectly respectable for a woman. This interpretation may seem to be a kind of shifting baselines, but we, at the very least, would argue that it's not simply an unconscious application of a different expectation: instead, people, as just discussed, consider it less respectable or even honorable for a woman to stay home to care for children. Once again, the fact that this effect diminishes in a different context provides further evidence. As with praiseworthiness, when the husband is paired with a wife who seems content with her decision to be a nurse (M=73.0, SD=18.4, 95%CI=[68.1-77.9]), he is no longer rated so much less intelligent than his counterpart, (M=75.2, SD=18.4, 95%CI=[70.5-79.9]). In fact, the effect disappeared: t(111)=0.670, p=0.504, d=0.125.

Dependent Variable: Decision-Making

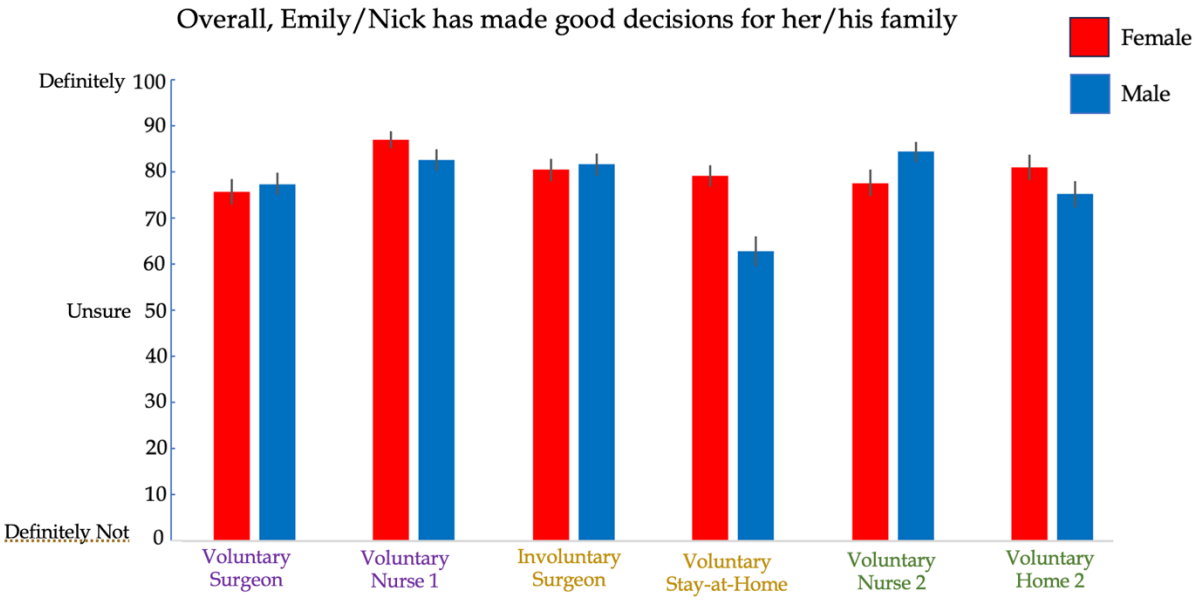

Figure 3.3. This depiction of "goodness of decisions" represents all six conditions and twelve contexts. Bars of the same color represent contexts that were combined into one condition. For instance, take the purple bars: the red, female voluntary surgeon was paired with the *blue,* male voluntary nurse in the conditions in Table 3.1. Errors bars represent *SE*.

A two-way ANOVA produced a main effect of gender, ($F(1, 517)=4.97$, $p=0.027$), a main effect of context ($F(5, 517)=9.56$, $p<0.001$), as well as an interaction $F(5, 517)=4.03$, $p<0.001$.

This measure provided strong support for our hypotheses. While the involuntary surgeon did not differ from its male counterpart, the female voluntary nurse with the ambitious surgeon husband ($M=87.1$, $SD=14.7$, 95%CI=[82.3-91.8]) was rated highest of all conditions once again. Though it did not differ from its male counterpart ($M=82.8$, $SD=18$, 95%CI=[79.1-86.5]; $t(112)=1.45$, $p=0.150$, $d=0.266$), it differed from all other male conditions, with the exception of the man who pursues nurse anesthesiology while his wife stays home with their child, $M=84.4$, $SD=17.2$, 95%CI=[80.0-88.7]. While these results might seem disparate, they both support our hypothesis that people endorse women who sacrifice for men and accept lesser public roles, responsibility, and prestige in comparison to their spouse. In fact, the interaction between gender

and one spouse staying at home while the other is a nurse anesthesiologist is just shy of significant at $F(1,112)=3.41$, $p=0.067$. Once again, the man who violates norms and chooses to stay home while his wife works the prestigious job of a surgeon is rated reliably lowest of all at $M=62.9$, $SD=17.7$, 95%CI=[58.3-67.5]. He is rated much lower than his female counterpart, ($M=62.9$, $SD=17.7$, 95%CI=[58.3-67.5]; $t(102)=4.18$, $p<0.001$, $d=0.789$. Interestingly, this female stay-at-home caregiver is about even with most of the female conditions, emphasizing just how strongly people value women in the home, especially when they are accompanied by accomplished men—like a surgeon in this case. Once again, this effect is attenuated in the comparison between the man and woman who stay home while their respective spouses work as nurses. Though in the predicted direction, the comparison is no longer significant: $t(112)=1.50$, $p=0.137$, $d=0.280$.

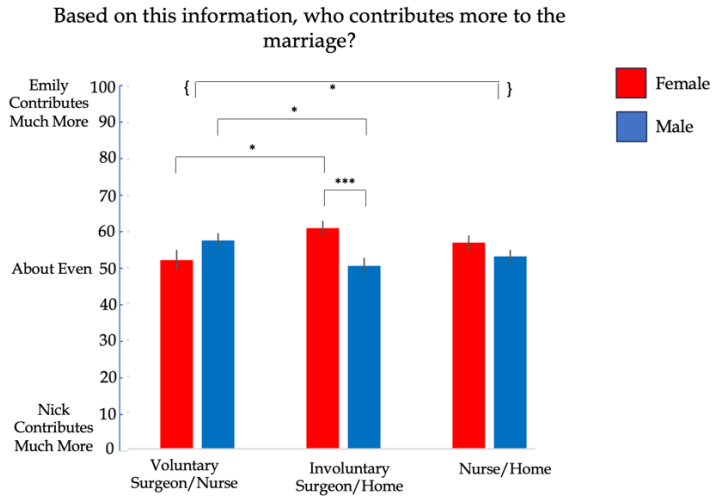

Figure 3.4. For this measure—and each of the following below—we simply compared the six conditions. The bars refer to whichever of the half of the condition is listed first. So, for instance, the first red bar refers to the female voluntary surgeon accompanied by the male voluntary nurse. The blue bar next to it represents the male voluntary surgeon accompanied by the female nurse. Error bars indicate *SE's*.

Dependent Variable: Contribution to the Marriage

The adjusted coefficient of determination for the model was R^2=0.030, F(5, 340)=3.81, p=0.008. There was a main effect of gender, (F(1, 340)=11.3, p<0.001, η_p^2= 0.032), a main effect of context, (F(2, 340)=4.13, p=0.017, η_p^2= 0.024), and an interaction between the two, (F(2, 340)=6.62, p=0.002, η_p^2= 0037).

As predicted, the context in which the female was an involuntary surgeon while her husband stayed at home with their child, was rated highest—i.e., that Emily contributed most to the marriage: M=61, SD=15.1, 95%CI=[57.1-64.9]. This condition differed substantially from its counterpart: M=50.5, SD=16.1, 95%CI=[46.3-54.7]; t(111)=3.56, p<0.001, d=0.670. This measure also clearly captured one of our predictions that was less clear for other DVs: that the woman who dares selfishly throws off social norms and takes a time-consuming, lucrative role more prestigious than her husband's—i.e., the voluntary surgeon—was rated much lower, (M=50.5, SD=16.1, 95%CI=[46.3-54.7) in the direction of Nick contributing more to the marriage. As might be expected, it differed from the involuntary surgeon, t(105)=2.60, p=0.011, d=0.482. Notably, the opposite was true for the inverse pairing: the voluntary male surgeon, (M=57.4, SD=16.4, 95%CI=[53.3-61.5]) was regarded as contributing more to Emily than the involuntary surgeon, t(115)=2.29, p=0.024, d=0.473. While either Nick's or Emily's actions could be responsible for a given rating, this pattern argues strongly for the interpretation that people are endorsing Emily's behavior as a woman compelled to take on a difficult, unsuitable role as a surgeon and also when she takes on the role of being a nurse to support an ambitious husband and have more time with her child.

Further, while neither pair-wise comparison was significant on its own, there was a significant interaction between the voluntary surgeons (paired with nurses) and the volunteering

nurses (paired with volunteering stay-at homes), that at least argued there was a different pattern for the respective pairings: $F(1, 222)=4.05$, $p=0.045$.

Dependent Variable: Financial Responsibility

The following model had an (adjusted) coefficient of determination at $R^2= 0.454$: $F(5, 240)=13.6$, $p<0.001$. There was a main effect of gender, ($F(1, 340)=11.3$, $p<0.001$, $\eta_p^2= 0.111$), a marginal effect of context, ($F(1, 340)=2.98$, $p=0.052$, $\eta_p^2= 0.017$), and an interaction between the two, ($F(2, 340)=5.08$, $p=0.007$, $\eta_p^2= 0.029$).

As expected, the results were skewed towards the men, ($M=57.4$, $SD=20.0$), having greater financial responsibility than women, ($M=52.6$, $SD=24.2$; $t(327)=7.49$, $p<0.001$, $d=0.809$). What's perhaps surprising is that this conception of male's greater financial responsibility should vary by condition, considering it was framed generically as "Who has the most financial responsibility: the husband or the wife? It's, of course, possible that people were thinking specifically about Nick and Emily.

All conditions were above the midline in the direction of Nick having greater responsibility, except for, unsurprisingly, the female involuntary condition (accompanied by husband nurse) who was at just about the midline: ($M=47.3$ $SD=24.6$, 95%CI=[40.9-53.7]). It appears that results were primarily dictated by whomever had the more public, prestigious role. When men had this position, the condition was rated higher towards Nick's responsibility and when women had this position, the opposite was true, relatively speaking, as men had a much higher baseline. Intriguingly, the two involuntary surgeons displayed the greatest disparity of the pairs, as well as highest and lowest among all conditions. The male, involuntary surgeon produced the following result: ($M=74.2$, $SD=10.0$, 95%CI=[69.0-79.4]). As is evident in Figure 3.5A below, the two conditions differed substantially: $t(105)=6.38$, $p<0.001$, $d=1.20$.

Interestingly, for this measure, both male and female involuntary surgeons differed from their voluntary counterparts. In each case, it seems, that pressure from one's spouse prompted the same decision to be perceived as imbibing more responsibility. The female involuntary surgeon was skewed more towards Emily than the voluntary surgeon condition, (M=57.3, SD=20, 95%CI=[52.1-62.5); t(111)=2.35, p=0.021, d=0.441. Likewise, the same was true of their male counterparts: In the context of the male voluntary surgeon, men were deemed to have less financial responsibility than in the context of the involuntary male: (M=65.9, SD=17.7, 95%CI=[61.4-70.4]); t(111)=2.40, p=0.018, d=0.444.

The nurse and home conditions were also dramatically different. In the context of the man working, (M=71.8, SD=10.0, 95%CI=[69.0-79.4]), husbands were rated as having greater financial responsibility than its counterpart condition: (M=53.8, SD=21.4, 95%CI=[48.3-59.2]), t(104)=4.09, p<0.001, d=0.773. While people clearly regarded men as being overall more responsible for finances, introducing different contexts changed perceptions of responsibility.

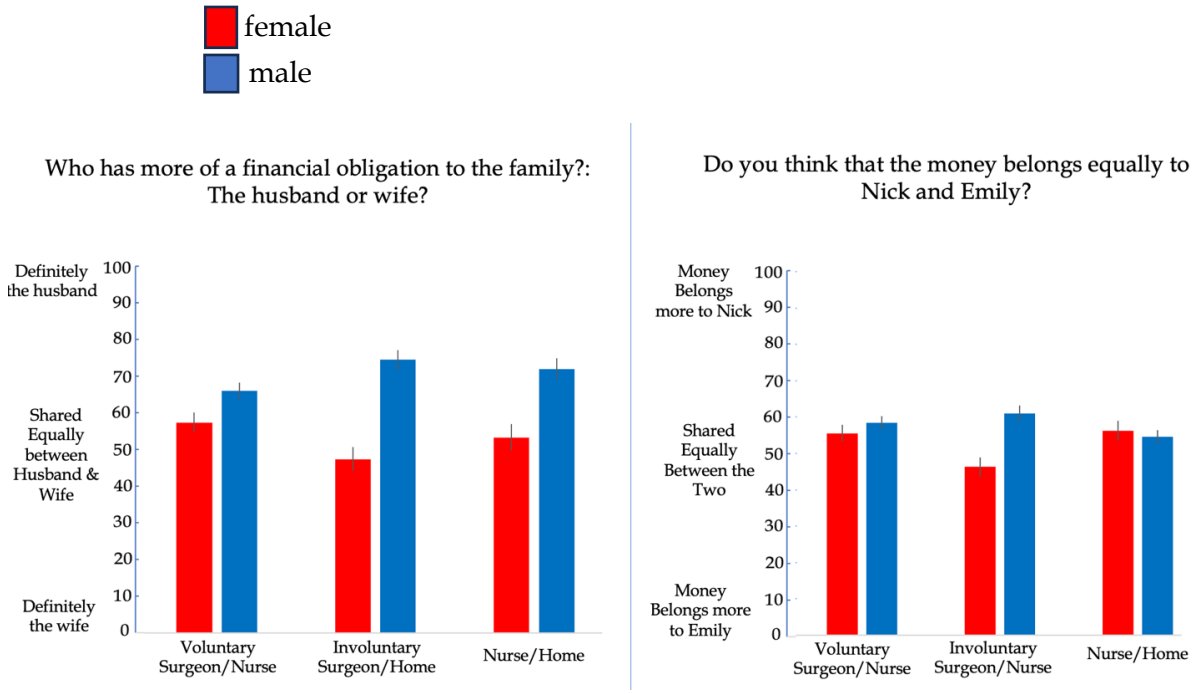

Figure 3.5. A) Represents general financial responsibility and B) represents the question of which spouse the money belongs to. For these measures we simply compared the six conditions. The bars refer to whichever of the half of the condition is listed first. So, for instance, the first red bar refers to the female voluntary surgeon accompanied by the male voluntary nurse. The blue bar next to it represents the male voluntary surgeon accompanied by the female nurse. Error bars indicate *SE's*.

Dependent Variable: Money

The model had an (adjusted) coefficient of determination at R^2= 0.051, $F(5, 240)=4.74$, $p<0.001$. There was a main effect of gender, ($F(1, 340)=20.5$, $p<0.001$, $\eta_p^2= 0.057$), a marginal effect of context, ($F(1, 340)=5.84$, $p=0.003$, $\eta_p^2= 0.033$), and an interaction between the two, ($F(2, 340)=6.76$, $p=0.001$, $\eta_p^2= 0.038$).

Results were skewed towards men, (M=57.9, SD=15.2), who overall were deemed to have greater ownership of the family money than women, (M=52.7, SD=19.7; $t(318)=2.78$, $p=0.006$, $d=0.300$. Once again, the largest point of interest in these results is that the involuntary surgeons both differed from one another and were each rated most "highest," (M=60.9,

SD=16.4, 95%CI=[56.7-65.2]), in the direction of Nick and "lowest" in the direction of Emily, (*M*=46.2, *SD*=20.4, 95%CI=[41.9-50.5]; $t(105)=4.21$, $p<0.001$, $d=0.795$) respectively.

General Discussion

The combined results for these measures provided strong support for our main hypotheses. As perhaps one of the most consistent results, the nurse paired with a prestigious surgeon for a husband was rated highly most highly in praiseworthiness and extremely highly in decision-making, presumably because 1) as a nurse, she takes on a role that, in a worldly sense, has less prestige than her ambitious surgeon husband and 2) because she cites time with her child. She was rated more highly than the other nurse (with a stay-at-home husband), though only in magnitude, as the comparisons were just short of significance. Notably, these two conditions were almost even with regard to intelligence, which supports our interpretations: stated simply, people tend to fault women who have more worldly responsibility or prestige and they especially resent it when choice has been involved. A possible explanation for this low rating of choice is that choice implies power. As further support for these interpretations, the female involuntary surgeon was rated highest among the surgeons in praiseworthiness—though the comparison was only significant with the voluntary male surgeon. Evidently, people praised her more—or faulted her less—for quite not being able to exercise as much *choice.* The "contribution to marriage" measure provided clear support for this perspective. Simply the perception of having taken a prestigious career (while her husband stayed home) earned her the highest contribution to the marriage. She was wildly higher than her male counterpart and, importantly, the female surgeon, who had the exact same role in a different context, daring to be ambitious.

These points acquire greater significance with consideration of possibly the most dramatic effect. The man who decided to stay-at -home to care for his child, “pressuring” his wife to be a surgeon (and presumably depriving her of her natural role with their child) was judged very harshly across DVs: He was regarded as far less praiseworthy, as having made comparatively bad decisions, and even as being much less intelligent than not only his female counterpart but those in all other conditions. Most notably, he was rated much differently than the man who made the same decision for the same reason but in a different context: with a wife who wanted to work as a nurse but was not cited as being particularly ambitious. In general, the results provided support for our main hypotheses: that prestige and dominance over their spouses is regarded as the natural position of men while women are faulted for the same attributes: ambition, choice, and less time with children.

FINAL THOUGHTS

While Chapters 1, 2, and 3 explored different lines of research, they converge in the common investigation of how people perceive harm and threat—more particularly how onlookers responded to it under various dynamic circumstances involving relationships with close others and strangers. For instance, though retaliation may have a rather negative connotation, our studies strongly suggest that people endorse it when certain features are involved. As discussed, people may consider it the particular domain of allies, serving to support existing bonds—or at least as being distinct from the mechanism supporting third-party retaliation. Participant reaction to slurs provided strong evidence that misogyny is not considered nearly so pernicious as racism. Yet the combination of the two studies emphasizes that bigotry, to state the obvious, is complicated. Further research could take many directions for all three lines of research.

Yet, it is also worthwhile and interesting to consider the three in combination. For instance, while people clearly did not consider the exhibition of misogyny to be as bad as that of black racism in Chapter 2, Chapter 3 demonstrated that—as research has already started to show—discrimination is likely at least partially motivated by underlying fears about the ascendence of women into equal rights. Like the water surrounding Bourdieu's fish, it's a strange juxtaposition of circumstances: to simultaneously love and seek to undermine the freedom of the person closest to one, whether conscious or not. All three chapters, in one way or another, involve the regulation of relationships and the nurturing and also avoidance of exploitation within and among those relationships.

APPENDIX

TABLE A.1: Chapter 2, Study 1 model for each dependent measure

DEPENDENT MEASURE	*F*	*p*	R^2(adjusted)/η^2
Bigotry: Based on his behavior, is [Agent] a bigot?			
	$F(1,186)=24.4$	**$p<0.001$**	**$R^2= 0.271/\eta^2= 0.282$**
Racism:	M=69.9	SD=18.6	n=45 95%CI=[64.5 -75.3]
Hybrid:	M=60.8	SD=30.3	n=50 95%CI=[52.4-69.2]
Misogyny:	M=45.6	SD=30.5	n=47 95%CI=[48.3-54.3]
Control:	M=27.3	SD=21.8	n=47 95%CI=[21.1-33.5]
Morality: Agree or Disagree: [Agent] is likely a moral person overall.			
	$F(1,186)=9.80$	**$p<0.001$**	**$R^2=0.123/\eta^2= 0.136$**
Racism:	M=37.3	SD=22.6	n=45 95%CI=[30.7-43.9]
Hybrid:	M=48.2	SD=24.0	n=50 95%CI=[41.5-54.9]
Misogyny:	M=54.7	SD=22.3	n=47 95%CI=[48.3-61.1]
Control:	M=61.1	SD=19.1	n=47 95%CI=[55.8-66.5]
Goodness-of-Action: Was [Agent]'s behavior good or bad?			
	$F(3,186)=12.2$	**$p<0.001$**	**$R^2=0.151/\eta^2= 0.164$**
Racism:	M=13.3	**SD=15.2**	n=45 95%CI=[8.86-18.5]
Hybrid:	M=20.5	SD=22.2	n=50 95%CI=[14.0-26.6]
Misogyny:	M=21.9	SD=20.0	n=47 95%CI=[16.1-27.7]
Control:	M=36.4	SD=17.7	n=47 95%CI=[43.0-53.0]
Punishment/Praise: Should [Agent] be punished or praised for his behavior?			

TABLE A.1 (continued):

$F(3,186)=10.6$ $p<0.001$ $R^2=0.133/\eta^2=0.146$

Racism:	M=76.0	SD=16.2	n=45	95%CI=[71.3-80.7]
Hybrid:	M=70.1	SD=21.7	n=50	95%CI=[64.1-76.1]
Misogyny:	M=67.6	SD=17.6	n=47	95%CI=[62.6-72.6]
Control:	M=56.0	SD=13.7	n= 47	95%CI=[52.1-59.9]

Sexism: Based on his behavior, is [Agent] likely sexist?

$F(3,186)=12.1$ $p<0.001$ $R^2=0.150/\eta^2=0.163$

Racism:	M=50.3	SD=22.1	n=45	95%CI=[43.8-56.7]
Hybrid:	M=53.8	SD=24.0	n=50	95%CI=[45.7-61.9]
Misogyny:	M=55.2	SD=31.3	n=47	95%CI=[46.3-64.1]
Control:	M=26.6	SD=23.2	n=47	95%CI=[20.1-33.1]

Racism: Based on his behavior, is [Agent] likely racist?

$F(3,186)=33.5$ $p<0.001$ $R^2=0.340/\eta^2=0.351$

Racism:	M=72.8	SD=21.4	n=45	95%CI=[65.6-80.0]
Hybrid:	M=61.2	SD=31.1	n=50	95%CI=[52.6-69.8]
Misogyny:	M=33.9	SD=25.2	n=47	95%CI=[26.7-41.1]
Control:	M=28.2	SD=22.1	n=47	95%CI=[22.0-34.4]

Jerk: Is [Agent] a jerk?

$F(3,186)=14.7$ $p<0.001$ $R^2=0.178/\eta^2=0.191$

Racism:	M=73.6	SD=25.0	n=45	95%CI=[66.3-80.9]
Hybrid:	M=65.6	SD=28.0	n=50	95%CI=[57.9-73.4]
Misogyny:	M=57.6	SD=32.5	n=47	95%CI=[48.3-66.9]
Control:	M=37.4	SD=25.1	n=47	95%CI=[30.3-44.5]

TABLE A.1 (continued):

Friendship: Do you think you would be friends with [Agent]?

$F(3,186)=8.61$ $p<0.001$ $R^2=0.108/\eta^2=0.122$

Racism:	M=26.9	SD=26.1	n=45	95%CI=[19.3-34.5]
Hybrid:	M=35.3	SD=27.1	n=50	95%CI=[27.7-42.9]
Misogyny:	M=38.3	SD=28.4	n=47	95%CI=[30.2-46.4]
Control:	M=54.0	SD=24.1	n=47	95%CI=[-47.2-60.8]

Respect: Agree or Disagree: [Agent] is probably a closed-minded person who doesn't respect people who are different from him.

$F(3,186)=15.9$ $p<0.001$ $R^2=0.191/\eta^2=0.204$

Racism:	M=64.4	SD=25.0	n=45	95%CI=[57.1-71.7]
Hybrid:	M=53.9	SD=29.5	n=50	95%CI=[45.7-62.1]
Misogyny:	M=44.0	SD=29.5	n=47	95%CI=[36.7-55.3]
Control:	M=27.6	SD=22.7	n=47	95%CI=[21.2-34.0]

Aggression: Agree or Disagree:[Agent]'s actions were aggressive.

$F(3,186)=5.57$ $p=0.001$ $R^2=0.068/\eta^2=0.082$

Racism:	M=82.7	SD=13.4	n=45	95%CI=[78.8-86.6]
Hybrid:	M=78.0	SD=20.1	n=50	95%CI=[72.4-83.6]
Misogyny:	M=74.5	SD=24.7	n=47	95%CI=[67.5-81.5]
Control:	M=65.9	SD=22.6	n=47	95%CI=[59.5-72.3]

TABLE A.2: Study 1— Pair-wise comparisons for each of the 10 dependent variables

DEPENDENT MEASURE COMPARISON/mean	mean difference	*t*	*p*	*d*
Bigotry : the higher the mean rating, the *more* of a bigot the Agent in the respective condition,				
Racism (69.6) vs. Hybrid (60.8)	8.8[-1.0,19]	*t*(82.6)=1.78	0.078	0.358
Hybrid (60.8) vs. Misogyny (45.6)	15.[2.9, 27]	*t*(95)=2.50	0.015	0.501
Misogyny(45.6) vs. Control (29.1)	17[7.4,29]	*t*(83.3)=3.36	0.001	0.446
Racism vs. Misogyny	23[14,35]	*t*(76.7)=4.64	<0.001	0.959
Morality: the higher the mean, the *more* moral people rated the Agent.				
Racism (37.3) vs. Hybrid (48.2)	10[14, 35]	*t*(92.8)= 2.27	0.025	0.594
Hybrid (48.2) vs. Misogyny (54.7)	2.5[-2.0, 18]	*t*(95.0)=1.37	0.171	0.318
Misogyny(54.7) vs. Control (61.1)	6.4[-2.0,14]	*t*(95.3)=1.52	0.133	0.312
Racism vs. Misogyny	17[13, 33]	*t*(102)= 4.45	<0.001	0.892
Hybrid vs. Control	12.9[-22,-4.2]	*t*(92.8)= 4.45	0.004	0.595
Goodness-of-Action: the higher the mean, the better participants rated the Agent's behavior.				
Racism (13.3) vs. Hybrid (20.5)	7.2[-0.56,15]	*t*(87.1)= 1.84,	0.069	0.481
Hybrid (20.5) vs. Misogyny (21.9)	1.4[2.9,27]	*t*(85.6)=2.30	0.023	0.371
Misogyny(21.9) vs. Control (48.0)	26.1[7.9,24]	*t*(91.1)=3.75	<0.001	0.771
Racism vs. Misogyny	8.6[1.2,16]	*t*(85.6)= 2.31	0.023	0.479
Punishment/Praise: the higher the mean, the more participants deemed the Agent worthy of punishment				
Racism (76.0) vs. Hybrid (70.1)	7.2[-1.9,14]	*t*(89.9)= 1.84	0.136	0.231
Hybrid (70.1) vs. Misogyny (67.7)	2.4[-5.5,10]	*t*(93.0)=0.62	0.536	0.392
Misogyny(67.6) vs. Control (56.0)	11.6[-18,-5.2]	*t*(86.9)=3.58	<0.001	1.17
Racism vs. Misogyny	8.3[1.4,15]	*t*(89.9)=2.38	<0.020	0.138
Sexism: the higher the mean, the more participants rated the Agent as sexist				

TABLE A.2 (continued):

Racism (50.3) vs. Hybrid (53.8)	3.5[-1.9,14]	$t(90.5)= 0.661$,	0.510	0.134
Hybrid (53.9) vs. Misogyny (55.2)	1.3[10, 14]	$t(93.4)=0.223$	0.823	0.045
Misogyny(55.2) vs. Control (23.2)	32[20,41]	$t(91.0)=5.82$	<0.001	1.04
Racism vs. Misogyny	4.9[-6.3,16]	$t(83.1)=0.866$	0.389	0.179
Racism vs. Control	27[14,33]	$t(91.0)=5.82$	<0.001	1.05

Racism: the higher the mean, the more participants determined the Agent to be racist.

Racism (72.8) vs. Hybrid (61.3)	11.5[0.72, 22]	$t(87.4)= 2.12$,	0.037	0.245
Hybrid (61.3) vs. Misogyny (34.0)	27.3[16,39]	$t(93.0)=4.78$	<0.001	0.847
Misogyny(34.0) vs. Control (23.2)	10.8[-3.9,15]	$t(90.0)=1.18$	0.240	0.43
Racism vs. Misogyny	38.9[29,48]	$t(88.8)= 7.98$	<0.001	0.601

Jerk: the higher the rating, the more participants considered the Agent a jerk.

Racism (73.6) vs. Hybrid (65.6)	8.0[-2.9, 19]	$t(93.0)= 1.46$	0.148	0.298
Hybrid (65.6) vs. Misogyny (57.6)	8.0[4.0,27]	$t(91.1)=1.30$	0.198	0.264
Misogyny(57.6) vs. Control (37.3)	20[-19,2.3]	$t(85.6)=3.39$	0.001	0.698
Racism vs. Misogyny	16[4.0,28]	$t(86.0)=2.65$	0.001	0.549

Friendship: the higher the rating, the more participants believed they would be friends with the Agent.

Racism (26.9) vs. Hybrid (35.3)	8.4[-2.4, 19.3]	$t(92.6)= 1.54$	0.125	0.317
Hybrid (35.3) vs. Misogyny (38.3)	3.0[-8.2, 14.1]	$t(89.9)=0.528$	0.599	0.107
Misogyny(38.3) vs. Control (54.0)	16[4.9-26.4]	$t(90.0)=3.39$	0.005	0.596
Racism vs. Misogyny	16[0.14,27]	$t(89.9)=2.01$	0.047	0.419

Aggression: the higher the mean, the more the participants considered the Agent's actions to be aggressive.

Racism (82.7) vs. Hybrid (78.0)	7.4[-2.2,12]	$t(86.0)= 1.36$	0.179	0.273
Hybrid (78.0) vs. Misogyny (75.4)	2.6[-5.6,13]	$t(88.9)=0.775$	0.441	0.158
Misogyny(74.5) vs. Control (65.9)	8.6[-1.0, 18]	$t(91.9)=1.77$	0.080	0.364

TABLE A.2 (continued):

Racism vs. Misogyny	7.3[0.06, 16]	$t(71.6)=2.01$	0.048	0.414
Hybrid vs. Control	12[3.6,21]	$t(93.7)=2.81$	0.006	0.569

Respect: the higher the mean, the less participants believed that the Agent respects people who are different from him.

Racism (64.4) vs. Hybrid (53.9)	11[-0.65, 21]	$t(92.7)= 1.87$	0.064	0.227
Hybrid (53.9) vs. Misogyny (44)	9.9[-2.0,22]	$t(94.6)=1.65$	0.103	0.334
Misogyny(44) vs. Control (27.6)	16[5.7,27]	$t(86.3)=3.04$	0.003	0.625
Racism vs. Misogyny	20[9.0, 32]	$t(88.7)=3.57$	<0.001	0.742

TABLE A.3: Study 2 model for each dependent measure

DEPENDENT MEASURE	F	p	R^2(adjusted)/η^2
Bigotry: Based on his behavior, is [Agent] a bigot?			
	$F(3, 194)=16.0$	**$p<0.001$**	**$R^2=0.156/\eta^2=0.198$**
Racism:	$M=70.9$	$SD=25.3$	$n=48$ 95%CI=[63.4-77.8]
Hybrid:	$M=74.7$	$SD=19.8$	$n=48$ 95%CI=[67.5-81.3]
Misogyny:	$M=62.6$	$SD=24.2$	$n=52$ 95%CI=[56.0-68.2]
Control:	$M=44.6$	$SD=24.1$	$n=50$ 95%CI=[37.9-51.3]
Morality: Agree or Disagree: [Agent] is likely a moral person overall.			
	$F(3,194)=2.97$	**$p=0.03$**	**$R^2=0.033/\eta^2= 0.044$**
Racism:	$M=59.4$	$SD=28.0$	$n=48$ 95%CI=[51.5-67.3]
Hybrid:	$M=47.3$	$SD=26.3$	$n=48$ 95%CI=[39.8-54.8]
Misogyny:	$M=55.6$	$SD=25.2$	$n=52$ 95%CI=[48.7-62.4]
Control:	$M=61.3$	$SD=19.6$	$n=50$ 95%CI=[55.9-66.7]

TABLE A.3 (continued):

Goodness-of-Action: Was [Agent]'s behavior good or bad?

$F(3,194)=2.14$ $p=0.09$ $R^2=0.0168/\eta^2=0.0628$

Racism:	$M=47.5$	SD=34.5	$n=48$	95%CI=[37.9-57.2]
Hybrid:	$M=31.2$	$SD=22.2$	$n=48$	95%CI=[14.0-26.6]
Misogyny:	$M=42.9$	$SD=33.5$	$n=52$	95%CI=[33.8-52.0]
Control:	$M=42.8$	$SD=27.9$	$n=50$	95%CI=[36.1-49.3]

Punishment/Praise: Should [Agent] be punished or praised for his behavior?

$F(3,194)=10.1$ $p<0.001$ $R^2=0.301/\eta^2=0.135$

Racism:	$M=74.4$	$SD=17.1$	$n=48$	95%CI=[69.6-79.2]
Hybrid:	$M=78.2$	$SD=15.1$	$n=48$	95%CI=73.9-82.5]
Misogyny:	$M=72.1$	$SD=15.2$	$n=52$	95%CI=[68.0-76.2]
Control:	$M=61.0$	$SD=31.4$	$n=50$	95%CI=[52.6-69.4]

Sexism: Based on his behavior, is [Agent] likely sexist?

$F(3, 194)=17.2$ $p<0.001$ $R^2=0.198/\eta^2=0.210$

Racism:	$M=59.0$	$SD=27.6$	$n=48$	95%CI=[51.2-66.8]
Hybrid:	$M=73.4$	$SD=20.6$	$n=48$	95%CI=[67.6-79.2]
Misogyny:	$M=73.6$	$SD=21.0$	$n=52$	95%CI=[67.9-79.3]
Control:	$M=43.3$	$SD=24.8$	$n=50$	95%CI=[36.4-50.1]

Racism: Based on his behavior, is [Agent] likely racist?

$F(3,194)=33.5$ $p<0.001$ $R^2=0.340/\eta^2=0.351$

Racism:	$M=71.5$	$SD=23.5$	$n=48$	95%CI=[64.9-78.1]
Hybrid:	$M=77.0$	$SD=21.2$	$n=48$	95%CI=[71.02-83.0]

TABLE A.3 (continued):

Misogyny:	*M*=56.1	*SD*=27.6	*n*=52	95%CI=[48.6-63.6]
Control:	*M*=40	*SD*=22.1	*n*=50	95%CI=[22.0-34.4]

Jerk: Is [Agent] a jerk?

$F(3,194)=25.7$ $p<0.001$ $R^2=0.274/\eta^2=0.285$

Racism:	*M*=74.1	*SD*=24.3	*n*=48	95%CI=[67.2-80.4]
Hybrid:	*M*=77.3	*SD*=18.1	*n*=58	95%CI=[72.2-82.4]
Misogyny:	*M*=71.6	*SD*=23.2	*n*=52	95%CI=[65.3-77.9]
Control:	*M*=40.6	*SD*=27.9	*n*=50	95%CI=[32.9-48.3]

Friendship: Do you think you would be friends with [Agent]?

$F(3,194)=2.31$ $p=0.077$ $R^2=0.020/\eta^2=0.035$

Racism:	*M*=51.1	*SD*=33.7	*n*=48	95%CI=[41.6-60.6]
Hybrid:	*M*=38.1	*SD*=32.8	*n*=48	95%CI=[28.8-47.4]
Misogyny:	*M*=47.2	*SD*=31.8	*n*=52	95%CI=[38.6-56.7]
Control:	*M*=53.8	*SD*=22.5	*n*=50	95%CI=[47.6-60.0]

Respect: Agree or Disagree: [Agent] is probably a closed-minded person who doesn't respect people who are different from him.

$F(3, 194)=10.0$ $p<0.001$ $R^2=0.121/\eta^2=0.134$

Racism:	*M*=67.8	*SD*=26.7	*n*=48	95%CI=[-60.3-75.3]
Hybrid:	*M*=73.3	*SD*=21.2	*n*=48	95%CI=[65.7-80.9]
Misogyny:	*M*=70.0	*SD*=28.4	*n*=52	95%CI=[62.4-77.5]
Control:	*M*=47.3	*SD*=22.3	*n*=50	95%CI=[39.7-54.8]

Aggression: Agree or Disagree:[Agent]'s actions were aggressive.

$F(3,186)=5.57$ $p=0.001$ $R^2=0.068/\eta^2=0.082$

TABLE A.3 (continued):

Racism:	*M*=79.6	*SD*=19.5	*n*=48	95%CI=[75.7-83.5]
Hybrid:	*M*=81.4	*SD*=15.5	*n*=48	95%CI=[77.0-85.8]
Misogyny:	*M*=77.3	*SD*=20.2	*n*=52	95%CI=[71.8-82.7]
Control:	*M*=72.9	*SD*=18.8	*n*=50	95%CI=[67.7-78.0]

REFERENCES

Almy, A., & Sanatullova-Allison, E. (2016). Women are the breadwinners and men are the homemakers: Gender socialization in culture, society, and education. *IAFOR Journal of Cultural Studies*, *1*(1), 1-7.

Alon, S., & DiPrete, T. A. (2015). Gender differences in the formation of a field of study choice set. *Sociological science*, *2*, 50.

Anderson, E. (2000). *Code of the street: Decency, violence, and the moral life of the inner city*. W Norton & Company.

Anderson, A., Chilczuk, S., Nelson, K., Ruther, R., & Wall-Scheffler, C. (2023). The Myth of Man the Hunter: Women's contribution to the hunt across ethnographic contexts. *PLoS one*, *18*(6), e0287101.

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607-1636.

Balafoutas, L., & Nikiforakis, N. (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review*, *56*(8), 1773-1785.

Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences*, *111*(45), 15924-15927.

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59-78.

Boehm, C. (1987). *Blood revenge: The enactment and management of conflict in Montenegro and other tribal societies*. University of Pennsylvania Press.

Brezina, T., Agnew, R., Cullen, F. T., & Wright, J. P. (2004). The code of the street: A quantitative assessment of Elijah Anderson's subculture of violence thesis and its contribution to youth violence research. *Youth Violence and Juvenile Justice*, *2*(4), 303-328.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391.

Bian, L., Leslie, S. J., Murphy, M. C., & Cimpian, A. (2018). Messages about brilliance undermine women's interest in educational and professional opportunities. *Journal of Experimental Social Psychology*, *76*, 404-420.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*(4), 485.

Brezina, T., Agnew, R., Cullen, F. T., & Wright, J. P. (2004). The code of the street: A quantitative assessment of Elijah Anderson's subculture of violence thesis and its contribution to youth violence research. *Youth Violence and Juvenile Justice*, *2*(4), 303-328.

Burt, R. S., & Knez, M. (1995). Kinds of third-party effects on trust. *Rationality and society*, *7*(3), 255-292.

Cross, E. J., Overall, N. C., Low, R. S., & McNulty, J. K. (2019). An interdependence account of sexism and power: Men's hostile sexism, biased perceptions of low power, and relationship aggression. *Journal of Personality and Social psychology*, *117*(2), 338.

Carlsmith, K. M., & Darley, J. M. (2008). Psychological aspects of retributive justice. *Advances in experimental social psychology*, *40*, 193-236.

Carlsmith, K. M., Wilson, T. D., & Gilbert, D. T. (2008). The paradoxical consequences of revenge. *Journal of personality and social psychology*, *95*(6), 1316.

Camerer, C. F. (2010). Behavioural game theory. In *Behavioural and Experimental Economics* (pp. 42-50). Palgrave Macmillan, London.

Campbell, A., Shirley, L., Heywood, C., & Crook, C. (2000). Infants' visual preference for sex-congruent babies, children, toys and activities: A longitudinal study. *British Journal of Developmental Psychology*, *18*(4), 479-498.

Chen, H., Zeng, Z., & Ma, J. (2020). The source of punishment matters: Third-party punishment restrains observers from selfish behaviors better than does second-party punishment by shaping norm perceptions. *PloS one*, *15*(3), e0229510.

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, *373*(6511), 209-216.

Coleman, J. M., & Franiuk, R. (2011). Perceptions of mothers and fathers who take temporary work leave. *Sex Roles*, *64*, 311-323.

Cushman, F., Gray K., Gaffrey, A. & Mended, W.B. (2012). Stimulating murder: the aversion to harmful action. *Emotion, 12*(1), 2.

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*(2), 281-299.

Dickerson, A., & Taylor, M. A. (2000). Self-limiting behavior in women: Self-esteem and self-efficacy as predictors. *Group & Organization Management*, *25*(2), 191-210.

Dreber, A., Rand, D. G., Fudenberg, D., & Nowak, M. A. (2008). Winners don't punish. *Nature*, *452*(7185), 348-351.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The developmental social psychology of gender*, *12*(174), 9781410605245-12.

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European review of social psychology*, *5*(1), 1-35.

Elshout, M., Nelissen, R. M., & van Beest, I. (2017). Your act is worse than mine: Perception bias in revenge situations. *Aggressive behavior*, *43*(6), 553-557.

Ericksen, K. P., & Horton, H. (1992). " Blood Feuds": Cross-Cultural Variations in Kin Group Vengeance. *Behavior Science Research*, *26*(1-4), 57-85.

Everett, J. A., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of experimental social psychology*, *79*, 200-216.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature*, *13*(1), 1-25.

Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition*, *123*(3), 434-441.

FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J. J., & Phelps, E. A. (2014). Fairness violations elicit greater punishment on behalf of another than for oneself. *Nature communications*, *5*(1), 5306.

Gaunt, R. (2013). Breadwinning moms, caregiving dads: Double standard in social judgments of gender norm violators. *Journal of family issues*, *34*(1), 3-24.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American psychologist*, *56*(2), 109.

Glick, P., & Fiske, S. T. (2018). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. In *Social cognition* (pp. 116-160). Routledge.

Gollwitzer, M., & Denzler, M. (2009). What makes revenge sweet: Seeing the offender suffer or delivering a message?. *Journal of Experimental Social Psychology*, *45*(4), 840-844.

Goeschl, T., & Jarke, J. (2016). Second and third-party punishment under costly monitoring. *Journal of Economic Psychology*, *54*, 124-133.

Grogan, S. K., (1992). Charles Fourier and the Nature of Women. *French Socialism and Sexual Difference: Women and the New Society, 1803–44*, 20-41

Gordon, D. S., Madden, J. R., & Lea, S. E. (2014). Both loved and feared: Third party punishers are viewed as formidable and likeable, but these reputational benefits may only be open to dominant individuals. *PLoS One*, *9*(10), e110045.

Günther, C., Ekinci, N. A., Schwieren, C., & Strobel, M. (2010). Women can't jump?—An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, *75*(3), 395-401.

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*(3), 417-449.

Govier, T. (2011). *Forgiveness and revenge*. Routledge.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*(1), 4.

Haidt, J., & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European journal of social psychology*, *26*(2), 201-218.

Haidt, J., Sabini, J., Gromet, D., & Darley, J. (2010). What exactly makes revenge sweet? How anger is satisfied in real life and at the movies. *Emotion (Under review), University of Virginia*.

Harinck, F., Shafa, S., Ellemers, N., & Beersma, B. (2013). The good news about honor culture: The preference for cooperative conflict management in the absence of insults. *Negotiation and Conflict Management Research*, *6*(2), 67-78.

Halevy, N., & Halali, E. (2015). Selfish third parties act as peacemakers by transforming conflicts and promoting cooperation. *Proceedings of the National Academy of Sciences*, *112*(22), 6937-6942.

Hughes, J., Creech, J. L., & Strosser, G. L. (2016). Attributions about morally unreliable characters: Relationship closeness affects moral judgments. *Basic and Applied Social Psychology*, *38*(4), 173-184.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, *60*(6), 581.

Jackson, J. C., Choi, V. K., & Gelfand, M. J. (2019). Revenge: A multilevel review and synthesis. *Annual Review of Psychology*, *70*, 319-345.

Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*, *113*(31), 8658-8663.

Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, *19*(12), 1280-1286.

Krasnow, M. M., Cosmides, L., Pedersen, E. J., & Tooby, J. (2012). What are punishment and reputation for?.

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human behavior*, *28*(2), 75-84.

Lieberman, D., & Linke, L. (2007). The effect of social category on third party punishment. *Evolutionary Psychology*, *5*(2), 147470490700500203.

Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford University Press.

Marshall, J., Mermin-Bunnell, K., & Bloom, P. (2020). Developing judgments about peers' obligation to intervene. *Cognition*, *201*, 104215.

Martin, A. E., & Mason, M. F. (2022). What does it mean to be (seen as) human? The importance of gender in humanization. *Journal of Personality and Social Psychology*, *123*(2), 292

McManus, R. M., Kleiman-Weiner, M., & Young, L. (2020). What we owe to family: The impact of special obligations on moral judgment. *Psychological Science*, *31*(3), 227-242.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, *36*(1), 1-15.

Mende-Siedlecki, P., Qu-Lee, J., Backer, R., & Van Bavel, J. J. (2019). Perceptual contributions to racial bias in pain recognition. *Journal of Experimental Psychology: General*, *148*(5), 863.

Meyer, M. A. (1989). Anti-semitism and Jewish identity. *Commentary*, *88*(5), 35-40.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of abnormal and social psychology*, *67*(4), 371.

Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of applied psychology*, *92*(4), 1159.

Momennejad, I., Sinclair, S., & Cikara, M. (2019). Computational justice: simulating structural bias and interventions. *BioRxiv*, 776211

Nelissen, R. M. (2008). The price you pay: Cost-dependent reputation effects of altruistic punishment. *Evolution and Human Behavior*, *29*(4), 242-248.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291-1298.

Perry, G. (2013). *Behind the shock machine: The untold story of the notorious Milgram psychology experiments*. New Press, The.

Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PloS one*, *9*(2), e88534.

Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in ecology & evolution*, *30*(2), 98-103.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological review*, *118*(1), 57.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: the hidden costs to women of a kinder, gentler image of middle managers. *Journal of personality and social psychology*, *77*(5), 1004.

Scaptura, M. N. (2019). *Masculinity threat, misogyny, and the celebration of violence in White men* (Doctoral dissertation, Virginia Tech)

Schlosser, A. E., & Levy, E. (2016). Helping others or oneself: How direction of comparison affects prosocial behavior. *Journal of Consumer Psychology*, *26*(4), 461-473.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106*(35), 15073-15078.

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2018). Being biased against friends to appear unbiased. *Journal of Experimental Social Psychology*, *78*, 104-115.

Shaw, A., & Knobe, J. (2013). Not all mutualism is fair, and not all fairness is mutualistic. *Behavioral and Brain Sciences*, *36*(1), 100.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological science*, *10*(1), 80-83.

Siy, J. O., & Cheryan, S. (2016). Prejudice masquerading as praise: The negative echo of positive stereotypes. *Personality and Social Psychology Bulletin*, *42*(7), 941-954.

Stillwell, A. M., Baumeister, R. F., & Del Priore, R. E. (2008). We're all victims here: Toward a psychology of revenge. *Basic and Applied Social Psychology*, *30*(3), 253-263.

Shteynberg, G., Gelfand, M. J., & Kim, K. (2009). Peering into the “magnum mysterium” of culture: The explanatory power of descriptive norms. *Journal of Cross-Cultural Psychology*, *40*(1), 46-69.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, *35*(1), 4-28.

Smith, J. L., Lewis, K. L., Hawthorne, L., & Hodges, S. D. (2013). When trying hard isn’t natural: Women’s belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Personality and Social Psychology Bulletin*, *39*(2), 131-143.

Tripp, T. M., & Bies, R. J. (2009). “Righteous” anger and revenge in the workplace: The fantasies, the feuds, the forgiveness. In *International handbook of anger: Constituent and concomitant biological, psychological, and social processes* (pp. 413-431). New York, NY: Springer New York.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, *46*(1), 35-57.

Upson, S., & Friedman, L. F. (2012). Where are all the female geniuses?. *Scientific American Mind*, *23*(5), 63-65.

Umera-Okeke, N. (2012). Linguistic sexism: an overview of the English language in everyday discourse. *AFRREV LALIGENS: An international journal of language, literature and gender studies*, *1*(1), 1-17.

Van Veelen, R., Derks, B., & Endedijk, M. D. (2019). Double trouble: How being outnumbered and negatively stereotyped threatens career outcomes of women in STEM. *Frontiers in psychology*, *10*, 150.

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and human behavior*, *32*(4), 231-262.

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027-1033.

Wilson, D. S., & Sober, E. (1998). Unto Others. *Harvard University Press, Carnbridge*, *9*, 95-112.

de Waal, F. B. M. (1996). *Conflict as negotiation.* In W. C. McGrew, L. F. Marchant, & T. Nishida (Eds.), *Great ape societies* (p. 159–172). Cambridge University Press.

Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology, 45*(3), 688–701.