Gender Differences in Severity Evaluation and Punitive Action

Viktoria Maria Focken

s4256816

Department of Psychology, University of Groningen

PSB3E-BT15 Bachelor Thesis

Group 44

Supervisor Dr. Maja Graso

Second Evaluator Gemma Schino

In collaboration with D.A. Anton, H. de. Vries, S.Y. Gang, M. Rihtaric, S. D. Kapelle

July 07, 2023

Abstract

This thesis investigates gender differences in the evaluations of an ambiguous allegation. Based on the body of literature, gender differences, not just in participant gender, but also in victim and perpetrator gender are expected to be found.The study presents participants with two vignettes describing an allegation of sexism. Both vignettes feature different allocations of victim and perpetrator gender. Participants are asked to rate the severity of the allegation and punishment towards the perpetrator in these cases. Severity is conceptualised on three dimensions of evaluations of the problem behaviour being serious, major and significant. Punishment is conceptualised through calls for discipline and dismissal. The results of the study indicate that women tend to evaluate both severity and punishment more severely than men. Furthermore, the female victim case, in general, received higher severity ratings from all participants. Notably, women exhibited a stronger inclination to punish the female perpetrator while men showed more leniency. These findings reveal intriguing gender differences in the assessment of ambiguous cases of sexism. The study suggests that women, compared to men, tend to adopt a more stringent approach in evaluating such situations. These results contribute to a deeper understanding of how individuals perceive and respond tounverfiable allegations, shedding light on the potential influences of gender in shaping judgement and decision-making processes.

Introduction

People often feel compelled to express their opinion and weigh in on situations that do not directly concern themselves. Many like to follow legal cases such as Depp v. Heard online and enjoy discussing with peers on social media such as Instagram or Twitter. Even before the case ended, people showed support for their favoured celebrity. It caused public uproar, considering the context of the #MeToo movement which highlights gender based violence and harassment in many areas. As the case progressed over the span of multiple weeks, more information about the plaintiff and defendant surfaced. Subsequently, many spectators revised their initial opinion and reevaluated who to support.

While the present study, unlike in the example above, focuses on non-physical allegations of harm, it nevertheless highlights the importance of judgement in ambiguous situations in which sufficient information is crucial to decision making. Detrimental mistakes can be made when supporting the wrong person and persecuting an innocent person. There is ambiguity in situations in which no clear side can be identified as *right* or *wrong*. In "my word against yours" cases or "hearsay" statements, such as Depp v. Heard, only the directly involved actors know what happened (Graso et al., 2019). As external observers, people quickly fall into fallacious reasoning.

According to Graso et al. (2023), when the evidentiary standards of a case are not met, people tend to rely more heavily on their mental schemata (Reynolds et al, 2020; Gray and Wegner, 2009). One such example is that observers heavily rely on stereotypes to fill in the gaps in order to make a judgement (Davies & Beech, 2012). Another bias known to affect decision making is ingroup favouritism in which a third party observer exhibits support for the person of their own gender (Rudman & Goodwin, 2004).

These decision-making errors can have negative consequences. For instance, in situations in which the observer cannot be sure about who is the victim and who is the perpetrator, the question remains how people make a judgement when there are consequences which require punitive action for the perpetrator. Therefore, the focus of this thesis is on exploring gender differences in judging ambiguous cases of harm concerning severity evaluations and punishment. An ambiguous "my word against yours" case is constructed in which the observer has only limited information about the true nature of the incident, there is no evidence.

Theoretical Background

This study focuses on gender differences in evaluations and decision-making in ambiguous situations. Findings from previous studies are examined in order to build a foundation for the hypotheses. First, I will review the literature on general gender differences to show why men and women might differ in their evaluations of harmful behaviour. This will be followed by gender stereotypes that affect evaluations of victims and perpetrators. Lastly, gender bias in decision making will be examined.

Gender Differences in Decision-Making

The first question is how men and women differ in their evaluations of harmful situations. Interestingly, the overall ratings of weakness and strength, victimisation or perpetration by male and female participants are considerably similar (Bracci et al., 2021). However, women seem to judge harmful behaviour as more severe than men do (Baron et al., 1991; Madan & Nalla, 2016; Pozzulo et al., 2009). Madan and Nalla (2016) examined gender differences in perceived seriousness and victimisation of sexual harassment. The results of their study show that women judged all harmful behaviours, non physical and physical, to be more serious than men did, resulting in significant mean differences between their judgments.

Behaviours such as unwanted sexual teasing, jokes, remarks or comments were perceived to be very serious by women and slightly less serious by men with a mean difference of 0.33 (Madan & Nalla, 2016). Other studies have found a similar mean difference in severity judgments at around 0.30, with women judging the severity of harassment as more serious (Rotundo et al., 2001). Baron et al. (1991) found that female participants rated sexist behaviour more severely than men did, especially coming from men. This more extreme tendency is also reflected by the research of Pozzulo et al. (2009), which revealed that female participants judged all victims to be more credible than male participants did. However, regarding punishing or sanctioning the perpetrator, male participants did react slightly more extreme than female participants (McKelvie, 2007). Other studies found no gender differences between punishment attitudes (Karlsson et al., 2021). The previously mentioned research evokes the question whether these differences also apply when the information at hand is less definitive, which led Allen and Nightingale (1997) to examine gender differences in ambiguous scenarios. In this case, the scenario depicted a child subjected to physical harm; however, due to the absence of evidence of harm, the testimony of the child was left uncorroborated (Allen & Nightingale, 1997). Despite the ambiguous circumstances, female participants tended to evaluate the allegation as more believable and important than male participants. Importantly, the scenario chosen for the study is an extreme and specific one, which raises the question whether this difference can also be found in other ambiguous contexts that do not focus on physical harm such as abuse. This question remains to be answered, since little research has been conducted on the gender differences in judging ambiguous, nonphysical harmful situations. This study aims to fill this gap in the research.

Based on the above findings, it is hypothesised that there is a significant difference in severity judgments between male and female participants, namely that female participants will express more severe opinions about supporting the victim as well as punishing the perpetrator.

H1: Female participants will evaluate an ambiguous allegation as more severe than men, and they will endorse stronger punishment than men.

Gender Stereotypes in Decision Making

Not only *who is judging* but also *who is being judged* is important when considering gender differences (Bracci et al., 2021; Hester & Gray, 2020). When it comes to evaluating ambiguous situations of harm and a judgement needs to be made, it must be decided who is the victim and who is the perpetrator to support the victim and call for due process of the perpetrator (Graso et al., 2023). Porter and tenBrinke (2009) found that in this process, initial impressions affect judgement heavily. According to Inman and Baron (1996) and Pozzulo et al. (2010), stereotypes are most active in ambiguous situations. Gender biases and stereotypes about men and women and victims and perpetrators can therefore have a notable impact on subsequent decision making. Women, for example, are strongly associated with the role of the victim, while men are assumed to have the role of the perpetrator (Reynolds et al., 2020). Indeed, in several studies it was found that the role of the victim was allocated to women, leading to them being given more support (Graso et al., 2023, Reynolds et al., 2020). According to Reynolds et al. (2020), this so-called typecasting happens when people make judgements based on certain mental templates that are based on stereotypes.

How come that men are typecast as the perpetrator and women as the victim? Firstly, this might be due to the fact that the female gender is associated with the well established stereotype of being victims of discrimination and harassment, not only in the workplace (Baron et al., 1991;

Burn, 2019; Castano et al., 2016; Parker & Funk, 2017). Secondly, men are associated with more aggression and violence as well as with competence and power, which is congruent with attributes, corresponding to the role of the perpetrator (Keough & Garcia, 2000; Reynolds et al., 2020; Richeson & Ambady, 2001; Rudman & Goodwin, 2004). Women, on the other hand, were associated with weakness and incompetence (Rudman & Goodwin, 2004). Over the course of the last two decades, gender stereotypes, especially those concerning women, have changed to be more positive (Charlesworth & Banaji, 2022; Eagly et al., 2020). Nevertheless, women continue to be viewed as weaker than men and weakness is associated with victims and victims are in need of support (Hine et al., 2022). Reynolds et al. (2020), propose that as a result, female employees would receive more support in ambiguous conflicts compared to men. Indeed, studies show that men are more inclined to typecast women as the victim, while women do not see other women in the role of the victim as much (Bracci et al., 2021; Glick & Fiske, 1996).

Female victims are clearly evaluated vastly different from male victims. In general, female victims were judged to be more credible than male ones (Voogt & Klettke, 2017). In contrast, male and female participants did not believe a male individual to have been a victim as much (Bracci et al., 2021).This is especially the case when the perpetrator is a woman (Davies et al., 2006; Pozzulo et al., 2010; Waldo et al., 1998).

As Reynolds et al. (2020) as well as Baron et al., in as early as 1991, found in their studies, female victims are more likely to be shown more support because they fit the stereotype of the victim while the male victim will not be shown the same amount of support. This poses the question of what happens when the perpetrator is female (Waldo et al., 1998). Interestingly, Reynolds et al. (2020) found that women, even when in the role of the perpetrator, were still punished less, with less severity and were “attributed qualities of victims” (Reynolds et al.,

2020). A potential explanation by Baron et al. (1991) implies that women are an unexpected source who are believed to be unable to discriminate against others, therefore they cannot be a perpetrator (Donnely & Kenyon, 1996). Interestingly, both male and female participants seem to engage in this typecasting behaviour. Baron et al. (1991) found that men and women equally judged the male perpetrator with more severity than the female perpetrator, even though committing the same action. Consequently, the female perpetrator will be treated more leniently[1] than the male perpetrator, who will receive greater punishment for the same offence (Baron et al., 1991).

H2: Observers will evaluate the female victim case as more severe in comparison to the male victim case and they will endorse greater punishment for the male perpetrator in comparison to the female perpetrator.

Gender Bias in Decision-Making

The last question I will evaluate is whether women will protect the female perpetrator in a way that they will not call for her punishment as much as they would for the male perpetrator. This is based on yet another bias that influences decision making processes in ambiguous situations, namely ingroup-bias (Rudman & Goodwin, 2004). Ingroup bias denotes the phenomenon of people of, for instance, the same gender showing more support for each other than for people of a different gender and even showing a dislike of members of the "outgroup" (Scheepers et al., 2006). Women are supposedly 4.5 times more likely to show an ingroup bias, compared to men who are supposedly more neutral and do not show such ingroup favouritism (Baron et al. 1991; Cappelen et al., 2017; Reynolds et al., 2020; Rudman & Goodwin, 2004). Accordingly, victims found more support from observers of the same gender (Inman, 2001). This does not only apply to victim support; female participants were also more lenient with the female

[1] Leniency refers to evaluating the allegation as less severe or being less stringent with the punishment

perpetrator regardless of the victim's gender, showing an ingroup bias (Meaux et al., 2018). In stark contrast, other researchers (Ahola et al., 2010) found a “same-sex penalty effect” in which the observer was less lenient with the perpetrator of their own gender and called for more severe punishment. Other studies found that male participants called for more severe punishment of the female perpetrator than for the male perpetrator (Meaux et al., 2018). Moreover, male participants sentenced the female perpetrator more severely when the victim was male, which does hint at an ingroup bias (Meaux et al., 2018). Based on this theory of ingroup favouritism, it is assumed that there is a significant gender difference in how men and women perceive the victim and perpetrator of their own gender. The fact that women seem to be more prone towards ingroup favouritism leads to the hypothesis that women will evaluate the female victim allegation with more severity and also be less severe with the punishment of the female perpetrator.

H3: There is a significant interaction of participant gender and victim/perpetrator gender, such that female participants will treat the female perpetrator less severely while male participants will not show this.

Present Study

With this study,the goal is to examine whether female participants do evaluate allegations as more severe than male ones, if the female victim is shown more support than the male victim and if the female perpetrator will be treated less severely by women and whether the male perpetrator will be treated more leniently by men. In the following sections I will explain the method, the results and discuss the findings with regards to implications as well as limitations. This study consists of a survey, asking participants to read two ambiguous cases which are “my word against yours” cases with different gender allocations in a victim/ perpetrator dyad. In

general, the question to be answered is “what influences people's perceptions of ambiguous situations?” Specifically, how does gender affect participants’ perception of an ambiguous, harmful allegation.

Method

Participants

Participants were recruited through various methods, including personal invitations to friends and family, as well as sharing a link of the survey in group chats predominantly consisting of psychology students. Efforts were made to mitigate convenience sampling by actively seeking participants outside of our immediate network. This was particularly important to ensure a diverse sample, including participants from different age groups. Furthermore, the study was also advertised on the Prolific platform, where 55 participants were recruited using a portion of the allocated research budget of 150€. Data collection was conducted using an online survey with XM Qualtrics Portal. Participants were provided with clear information about the study's purpose, their rights, and the confidentiality of their responses. Informed consent was obtained from all participants. The Ethics Committee approved of this study.

A total of 186 participants completed the survey, with 64% female and 36% male participants. The age range of participants spanned from under 20 to over 60 years, with the majority (38.9%) falling within the 20-25-year age group ($Mage$= 31.7).

Procedure

We created two nearly identical vignettes portraying an ambiguous allegation with one having a female victim and male perpetrator and the other one having a male victim and female perpetrator. In this case, the perpetrator, unlike committing a crime, said or did something allegedly harmful and the victim is the person that made the allegation. The vignettes are

displayed in Appendix A. The victim makes a claim against a colleague who allegedly told a tasteless joke. This claim is however uncorroborated as no other person heard it. The second case includes a counter argument with the different gender allocation. The issue of this vignette is that this situation is so vague, that third party observers cannot make a proper judgement based on the information given and have to rely on their biases, stereotypes and decision making heuristics in order to form a conclusion. This was done purposefully to identify what factors and potential biases influence the participant's evaluations given that greater ambiguity leads to greater stereotype activation (Inman & Baron, 1996). The sequence of the vignettes was randomised so that a proportion of the participants saw the female victim first, while others saw the male victim first. Importantly, this study focuses on female and male gender differences, participants with a different gender identity are therefore excluded from this research.

Measures

This study focuses on how the participants evaluate the severity of the allegation and how they call for punishment of the alleged perpetrator. Participants are administered the same set of questions, once for each victim case, to assess their evaluations of severity and punishment. The correlations between the variables of interest are significant, meaning they are associated with one another, see Appendix A, Table 1. The questions pertaining to the male victim and female victim cases exhibit positive correlation indicating a parallel evaluation pattern, where higher severity ratings for the victim are associated with higher severity ratings for the perpetrator.

Severity

The severity evaluation is operationalised through three different scales. Participants were asked to rate on a 7 point scale, how *serious- unserio*us, *major-minor* and

significant-insignificant they evaluated the allegation to be. The items are highly reliable with the female victim α = .897, male victim α = .916 and a combined reliability of α = .913.

Punitive Intentions

Punishment is conceptualised by asking the participant to indicate whether they would discipline and/ or dismiss the alleged perpetrator.

Call for discipline. Participants were asked whether they would discipline the perpetrator on a scale of 0 = *not at all* to 10 = *discipline severely*.

Dismissal. We asked participants to evaluate whether they would dismiss the perpetrator from the company, based on a zero tolerance policy. Zero tolerance policies are implemented to sanction the alleged perpetrator and leave no room to promote harmful behaviour (Stockdale et al., 2004). It is assumed that participants are aware of such a policy.

Results

Due to the nature of this study, results will be jointly reported to compare or contrast the gender differences. They will however be divided by the gender of the victim and will be interchangeably referred to as “female victim case” or “female victim allegation” and “male victim case” or “male victim allegation”. Participants are referred to by their gender as “female participants” or “women” and “male participants” or “men”. Results will be interpreted using the effect size Cohen’s d, provided in Cohen’s paper (1988). The data were analysed using IBM SPSS Statistics (Version 29).

Hypothesis 1: Gender Differences in Evaluating Severity and Punitive Action

For the first hypothesis, an independent sample *t* test was performed to determine whether there is a significant gender difference in the answers of men and women. Specifically,

whether female participants evaluated the vignette more severely than male participants. Because of this hypothesis, one sided p-values are examined at a significance level of $\alpha = .05$.

First, the assumptions were tested. The results are expected to be independent due to the sampling method. Normality was severely violated with the Shapiro Wilk test of $p < .001$ for every item. However, based on Normality QQ- Plots, the assumption is not as severely violated (see Appendix B Figure 1 - 4). The dismissal question is dichotomous, assessing its normality is therefore not necessary. For the assumption of homogeneity, a Levene's test was conducted to check for equal variances. The answers to the female victim "discipline" and "dismissal" question yielded significant Levene test results at $F = 10.632$, $p = .001$ and $F = 7.946$ and $p = .005$ respectively. The independent sample *t* test was conducted nevertheless, as it is a robust measure. For the above named variables, the corrected p-value will be assessed; for an easier overview, the significant Levene test variables will be marked with an asterisk.

The independent sample *t* test (see Table 1) yielded some significant results. Unless specified otherwise, the differences are in favour of the female participant severity evaluations. The result of the aggregated female victim severity scale shows a significant gender difference with women scoring higher than men. There was no significant gender effect for male victim severity, despite women attaining higher scores than men (see Table 2). The punishment evaluations show significant gender differences at $\alpha = 0.5$ for both victim gender cases with female participants scoring higher. The female victim discipline and the dismissal show gender differences with women's evaluations being higher than men's. The male victim punishment evaluations also differed in discipline and in dismissal opinion, however, with lower significance and a lower effect size (see Table 2). Interestingly, male participants called for the dismissal of the perpetrator, female and male, with the exact same amount of 45% voting for yes.

Table 1

Test of Gender Differences in Female Victim Evaluation

Female Victim	Gender	*M*	*SD*	*t*	*df*	*p*	d	Lower	Upper
Severity	Female	5.63	1.20	3.21	181	< .001	0.495	0.24	0.99
	Male	5.02	1.30						
Discipline*	Female	6.18	2.06	3.82	167	< .001	0.614	0.67	2.10
	Male	4.80	2.57						
Dismissal*	Female	0.68	0.47	3.02	182	.001	0.464	0.08	0.37
	Male	0.45	0.50						

Note. Different df reflect differences in the amount of answers because not every participant answered every question. The equal variances not assumed corrections are marked with an asterisk. All p-values are one-sided. The significance level is at alpha = 5%.

Table 2

Test of Gender Differences in Male Victim Evaluation

Male Victim	Gender	*M*	*SD*	*t*	*df*	*p*	d	Lower	Upper
Severity	Female	5.14	1.48	1.45	180	.0740	0.224	-0.12	0.76
	Male	4.81	1.38						
Discipline	Female	5.43	2.39	2.19	167	.0150	0.354	0.08	1.67
	Male	4.55	2.65						
Dismissal	Female	0.62	4.88	2.16	182	.0160	0.333	0.06	0.31
	Male	0.45	0.50						

Note. Different df reflect differences in answers because not every participant answered every question. All p-values are one-sided. The significance level is at alpha = 5%.

Hypothesis 2: Gender Stereotypes - Differences in Female and Male Victim Perceptions

For the second hypothesis, a paired sample *t* test was conducted to compare the evaluations of female vs. male victims as well as punishment of the female vs. male perpetrator. Whether the female victim's allegation is evaluated with more severity than the male victim's and whether male perpetrators are being punished more than female perpetrators. All pairs are significant at $\alpha = 0.05$. Mean differences and the respective effect sizes can be found in Table 3. For the severity evaluations there is a difference in favour of the female victim with more severe evaluations compared to the male victim. The discipline and dismissal mean differences are in favour of the female victim. Although yielding a significant result, dismissal has a strikingly small effect size. The evaluations are in favour of the female victim also shown by the greater punishment evaluations of the male perpetrator, although the difference is not as striking as for other variables.

Table 3

Test of Differences between Female and Male Victim and Perpetrator Evaluations

	M	*t*	*df*	*p*	d	Lower	Upper
FV Severity - MV Severity	0.343	4.23	193	< .001	0.307	0.19	0.50
FV Discipline - MV Discipline	0.490	4.41	170	< .001	0.337	0.27	0.71
FV Dismissal - MV Dismissal	0.036	1.95	194	.026	0.14	0.00	0.072

Note. Different df reflect differences in answers. Not every participant answered every question. FV denotes *female victim* and MV denotes *male victim*. Female victim has a male perpetrator and male victim has a female perpetrator.

Hypothesis 3: Gender Bias - Interaction of Observer Gender and Participant Gender

Lastly, to see for an effect of gender, a repeated measures analysis of variance (ANOVA) was conducted. The hypothesis poses the question of an effect between the gender of the observer on the gender of the victim or perpetrator with regards to their evaluations. It was assessed whether female participants show greater severity evaluations towards the female victim and more leniency towards the female perpetrator. The same question goes for the male participant, whether they show this bias towards the male victim and perpetrator. The normality assumption, as shown previously, is violated. Nevertheless, as seen in the Appendix B Figures 1-5, the QQ - Plots show that the assumption is not violated as much as predicted by the Shapiro Wilk statistic. The dismissal question, as it is dichotomous, cannot be assessed. The assumption of sphericity is met with Mauchly W = 1 for all items.

There is a significant main effect of severity, showing a change from female victim to male victim evaluations (see Table 5). The between group analysis shows a significant difference between men and women ($F = 5.792$, $p = .017$). However, there is no significant interaction between victim gender observer and gender on severity as seen in Table 5, but with $p = .053$ being just over the alpha level, it is worth keeping in mind. From the descriptive statistics, it can be seen that women did evaluate the female victim case with more severity while men did not show such bias (see Table 4).

In the next test, gender has a main effect at $F = 10.401$ with $p = .002$, showing difference between men and women as well as a significant main effect of discipline, showing a change from male perpetrator to female perpetrator evaluations (Table 7). There is also a significant interaction effect of observer gender and perpetrator gender on discipline evaluations as seen in Table 7. Women called for less discipline of the female perpetrator than for the male perpetrator

and men evaluated the male perpetrator more severely than the female one. Overall, women's evaluations were more severe as seen in Table 6.

In Table 9, the data does reveal a significant main effect of gender, showing a difference between men and women but no significant main effect for perpetrator gender dismissal nor an interaction between the perpetrator gender and observer gender on dismissal evaluation as seen in Table 9. This might be due to dismissal being a dichotomous variable which affects the repeated measures ANOVA. A different test would be more appropriate. In Table 8, it can be seen that women endorse the dismissal of the female perpetrator slightly more. However, as seen in the previous tests (see Table 1, 2 and 3) men and women called for dismissal with striking similarity, which is a more likely reason for the non-significance.

Table 4

Descriptive Statistics of Severity

	Gender	*M*	*SD*	*N*
Female Victim - Severity	Female	5.63	1.21	116
	Male	5.02	1.30	66
Male Victim - Severity	Female	5.14	1.48	116
	Male	4.82	1.39	66

Table 5

Analysis of Gender, Severity and Severity x Gender

	F	*p*	*ηp2*
Gender	5.792	.017	.031
Severity	21.945	< .001	.11
Severity x Gender	3.799	.053	.02

Note. df = 1. Severity is Female Victim Severity and Male Victim Severity

Table 6

Descriptive Statistics of Discipline

	Gender	*M*	*SD*	*N*
Female Victim - Discipline	Female	6.20	2.10	105
	Male	4.77	2.63	56
Male Victim - Discipline	Female	5.60	2.37	105
	Male	4.55	2.70	56

Note. Female Victim has a Male Perpetrator. Male Victim has a Female Perpetrator

Table 7

Analysis of Gender, Discipline and Discipline x Gender

	F	*p*	*ηp2*
Gender	10.401	.002	.061
Discipline	19.659	< .001	.069
Discipline x Gender	3.903	.05	.024

Note. df= 1. Discipline is Male Perpetrator Discipline and Female Perpetrator Discipline

Table 8

Descriptive Statistics of Dismissal

	Gender	*M*	*SD*	*N*
Female Victim - Dismissal	Female	0.68	0.47	118
	Male	0.45	0.50	66
Male Victim - Dismissal	Female	0.62	0.49	118
	Male	0.45	0.50	66

Note. Female Victim has a Male Perpetrator. Male Victim has a Female Perpetrator. Dismissal is a dichotomous variable,

Table 9

Analysis of Gender, Dismissal and Dismissal x Gender

	F	*p*	*ηp2*
Gender	7.138	.008	.038
Dismissal	2.561	.111	.014
Dismissal x Gender	2.561	.111	.014

Note. df = 1. Dismissal is not a continuous variable. Dismissal is Male Perpetrator Dismissal and Female Perpetrator Dismissal

Discussion

The findings can be divided into the three hypotheses. First the gender differences in severity evaluations, then the evaluations of female and male victims and perpetrators and lastly the interaction between observer gender and the gender allocations in the vignettes.

Based on the literature, it was hypothesised that women would evaluate harmful allegations as more severe than men would (Baron et al., 1991; Madan & Nalla, 2016; Pouzzulo et al., 2009). As anticipated, the data supports the previous research. The mean differences between women's and men's evaluations are around 0.3, similar to the ones Madan & Nalla (2016) and Rotundo et al. (2001) have found (see Table 1 and 2). This pattern was observed across multiple items of severity and punishment.

The second hypothesis was that the female victim case is evaluated with more severity than the male victim case and that subsequently the male perpetrator would be shown more punishment than the female perpetrator. Based on the literature, women are allocated the role of the victim and men the role of the perpetrator (Reynolds et al., 2020). The allegation of the female victim is indeed judged more severely which aligns with the research (Baron et al., 1991; Reynolds et al., 2020). The male perpetrator was also shown with more punitive intentions than the female perpetrator, as prior research predicted (Reynolds et al., 2020). In all cases, the evaluations were more in favour of the female victim.

The third hypothesis was that there would be an effect of observer gender and gender in the vignettes. On the basis of an ingroup bias, observers would evaluate the victim of their own gender with greater severity while concurrently also evaluating the perpetrator of their own gender with less severity. Regarding the interaction effect between victim/perpetrator gender and observer gender, the findings revealed some interesting patterns. As for punishment, there are some intriguing disparities. It was anticipated that, while men would not engage in such ingroup bias, that women would show greater leniency with the female perpetrator. When looking at the data of the independent sample *t* tests (see Table 1 and Table 2), it can be assumed that women show a same sex penalty effect in which the observer punishes the person of their own gender more (Ahola et al., 2010). This was observed as female participants called for more and harsher punishment of the female perpetrator. Remarkably, men called for far less punishment of the female perpetrator, much more so than women did. However, taking the evaluations of the male perpetrator into account, women were more lenient with the female perpetrator, alluding to an ingroup bias. Both men and women called for higher discipline of the male perpetrator.

Implications

The implications of these findings extend beyond individual perceptions and have broader implications for the legal and criminal justice system. The gender gap observed in the legal system, where female perpetrators are often handled with more leniency, may be influenced by stereotypes and biases that perceive women as incapable of discrimination or perpetration (Geppert, 2022). This bias is further reinforced by societal expectations and ingrained gender roles (Ellemers, 2018). Especially in ambiguous contexts in which decisions need to be made, these stereotypes and biases are dominant in influencing judgement (Inman & Baron, 1996).

Especially the legal system, it is important to make correct decisions even when there is little to no evidence. The people acting within the system must be able to accurately evaluate the credibility of the witness (Porter & ten Brinke, 2009; Rozmann & Levy, 2023). When it comes to being able to detect deception, people working in law enforcement have no advantage over ordinary third party observers (DePaulo & Pfeifer , 1986). Even judges show bias in their decision making (Porter & tenBrinke, 2009). Additionally, in the context of jury members, a balance of gender should be implemented as it might skew the judgments when one gender group is over- or underrepresented (Pozzulo et al., 2010). Because judges often encounter contradictory statements it is even more important to be aware of gender differences, biases, and stereotypes (Porter & ten Brinke, 2009). As seen by this study, gender differences alone can play a role in judging ambiguous cases of harm. By recognizing the potential for biassed judgments, policymakers, legal professionals, and organisations can work towards implementing measures that ensure equal treatment for all individuals, regardless of gender.

Limitations

While this study provides insight into the perceptions of harmful allegations, limitations must also be taken into consideration. These limitations include, but are not limited to the following. First, this study was developed and carried out as part of the bachelor psychology programme in collaboration with six other students. Because each one of us has a different approach to the overall research questions, multiple variables and questions were measured within the same design. This might have an influence on the answers of the participants as questions by other students might affect their answers to the questions that were examined in this present study. Second, due to time constraints, there is no pilot study which also affects the overall construction of the survey and mistakes that are noticed only afterwards. Third, although

efforts were made to ensure the validity of the measurement instruments used in this study, there is always a possibility of measurement error. Fourth, future studies should try to have coherent answer possibilities which don't affect the validity of the measurement instruments for higher statistical power. Fifth, the sample size which may limit the generalisability of the findings to a larger population. A larger sample size would enhance the statistical power and increase the confidence in the results. Next to the size, the sampling method also limits the generalizability. Although trying to achieve a broad sample population, it is not guaranteed that it is representative of the entire population which concludes a sixth limitation. Seventh, the length of the study is a limitation as the Qualtrics platform calculated the time needed to complete the survey at 17 min. Because surveys should be kept short to keep the attention of the observer, this also affected the vignettes as reading them takes more time and we could therefore not include as many. The eighth limitation is that we only have two vignettes; future research should include neutral cases to measure baseline answers. Ninth, the vignettes are supposed to elicit bias in the participants, however names of the characters with “Joan” and “Sam” are rather gender neutral. Future research should consider this and choose something that's more stereotypical of male and female gender. Tenth, in retrospect additional questionnaires about participants' stereotypes about victims and perpetrators should be included as there is currently not much research on these attitudes. It would have also been really interesting to analyse whether and how gender stereotypes play into this. Tenth, considering the context of this study, which is being conducted as part of my bachelor's degree, the choice of statistical tests is restricted. Because . Eleventh and last, measurements for response biases should be implemented to control for participants that are not answering truthfully.

Conclusion

I sought to examine gender differences in evaluations of ambiguous allegations of sexist behaviour. I was interested in not just the gender differences of the observers but also the complex gender differences in stereotypes and of biases. Despite the limitations acknowledged in this study, the findings do unveil intriguing patterns. Not just in the female victim but also in the male victim case, female participants evaluated the allegation as more severe. This underlines women's tendency towards evaluating harmful behaviour to be more stringent, not just in an ingroup context. Already in 1991, Baron et al. hypothesised that men and women would judge a male perpetrator more severely and this is supported by this present study. An interesting observation is that male participants called for the dismissal of the male and female perpetrator with exactly the same amount, namely 45%. The anticipated leniency towards the female perpetrator by female participants and lack of leniency towards the male perpetrator by male participants were not strongly evident. Although Glick & Fiske (1996) did find that men were more likely to typecast the female actor as the victim, it was believed that women's ingroup bias would have a greater effect than men's gender stereotypes (Rudman & Goodwin, 2004). However, these findings are evident of a same-sex penalty effect and call for more substantive research on this topic (Ahola et al., 2010). The findings underline the complex interplay of gender differences and individual perceptions such as gender stereotype and gender bias in evaluating ambiguous allegations. Further research is needed to explore the underlying factors and decision-making processes that influence these patterns of severity and punishment.

References

Ahola, A. S., Hellström, Å., & Christianson, S. Å. (2010). Is justice really blind? Effects of crime descriptions, defendant gender and appearance, and legal practitioner gender on sentences and defendant evaluations in a mock trial. *Psychiatry Psychology and Law, 17*(2), 304–324. doi:10.1080/13218710903566896

Allen, L. A., & Nightingale, N. N. (1997). Gender Differences in Perception and Verdict in Relation to Uncorroborated Testimony by a Child Victim. *Journal of Offender Rehabilitation, 24*(3-4), 101–116. doi:10.1300/j076v24n03_06

Baron, R. S., Burgess, M. L., & Kao, C. F. (1991). Detecting and labeling prejudice: Do female perpetrators go undetected? *Personality and Social Psychology Bulletin, 17*(2), 115–123. https://doi.org/10.1177/014616729101700201

Bracci, M., Guidi, S., Marchigiani, E., Masini, M., Palmitesta, P., & Parlangeli, O. (2021). Perception of Faces and Elaboration of Gender and Victim/Aggressor Stereotypes: The Influence of Internet Use and of the Perceiver's Personality. *Frontiers in psychology*, *12*, 561480. https://doi.org/10.3389/fpsyg.2021.561480

Burn, S. M. (2019). The Psychology of Sexual Harassment. *Teaching of Psychology, 46*(1), 96–103. https://doi.org/10.1177/0098628318816183

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Hillsdale, NJ: Lawrence Earlbaum Associates.

Davies, M., Pollard, P., & Archer, J. (2006). Effects of Perpetrator Gender and Victim Sexuality on Blame Toward Male Victims of Sexual Assault. *The Journal of Social Psychology, 146*(3), 275–291. doi:10.3200/socp.146.3.275-291

Davies, G. M., & Beech, A. R. (Eds.). (2012). *Forensic psychology: Crime, justice, law, interventions*. John Wiley & Sons.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-Job Experience and Skill at Detecting Deception1. *Journal of Applied Social Psychology, 16*(3), 249–267. doi:10.1111/j.1559-1816.1986.tb01138.x

Donnely, D. A., & Kenyon, S. (1996). "Honey, We Don't Do Men." *Journal of Interpersonal Violence, 11*(3), 441–448. doi:10.1177/088626096011003009

Ellemers, N. (2018). Gender stereotypes. *Annual review of psychology, 69*, 275-298.

Geppert, K. (2022). Explaining the Gender Gap in the Criminal Justice System: How Family-Based Gender Roles Shape Perceptions of Defendants in Criminal Court. *Inquiries Journal, 14*(02). Retrieved from http://www.inquiriesjournal.com/a?id=1936

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of personality and social psychology, 70*(3), 491.

Graso, Maja & Reynolds, Tania & Grover, Steven. (2019). Allegations of Mistreatment in an Era of Harm Avoidance: Taboos, Challenges, and Implications For Management. *Academy of Management Perspectives, 34*. 10.5465/amp.2018.0144.

Graso, M., Aquino, K., Chen, F. X., Camps, J., Strah, N., & van den Bos, K. (2023). When Do Observers Deprioritize Due Process for the Perpetrator and Prioritize Safety for the Victim in Response to Information-Poor Allegations of Harm? *Psychological Science, 34*(2), 186–200. https://doi.org/10.1177/09567976221128203

Gray, K., & Wegner, D. M. (2009). Moral typecasting: divergent perceptions of moral agents and moral patients. *Journal of personality and social psychology, 96*(3), 505–520. https://doi.org/10.1037/a0013748

Hester, N., & Gray, K. (2020). The Moral Psychology of Raceless, Genderless Strangers. *Perspectives on Psychological Science, 15*(2), 216–230. https://doi.org/10.1177/1745691619885840

Hine, B., Noku, L., Bates, E. A., & Jayes, K. (2022). But, Who Is the Victim Here? Exploring Judgments Toward Hypothetical Bidirectional Domestic Violence Scenarios. *Journal of Interpersonal Violence, 37*(7–8), NP5495–NP5516. https://doi-org.libproxy.abertay.ac.uk/10.1177/0886260520917508

Inman, M. L., & Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology, 70*(4), 727–739. https://doi.org/10.1037/0022-3514.70.4.727

Inman, M. L. (2001). Do You See What I See?: Similarities And Differences In Victims' And Observers' Perceptions Of Discrimination. *Social Cognition, 19*(5), 521–546. doi:10.1521/soco.19.5.521.19912

Karlsson, L. C., Malén, T., Kaakinen, J. K., & Antfolk, J. (2021). The Effect of Sex and Perpetrator–Victim Relationship on Perceptions of Domestic Homicide. *Journal of Interpersonal Violence, 36*(7–8), 3309–3330. https://doi-org.proxy-ub.rug.nl/10.1177/0886260518775162

Lawless, T., O'Dea, C., Miller, S. & Saucier, D. (2020). Is it really just a joke? Gender differences in perceptions of sexist humor. *Humor, 33*(2), 291-315. https://doi.org/10.1515/humor-2019-0033

Madan, M., & Nalla, M. K. (2016). Sexual harassment in public spaces: Examining gender differences in perceived seriousness and victimization. *International criminal justice review*, *26*(2), 80-97.

McKelvie, S. J. (2007). Some sex differences in sentencing severity for a mock crime scenario. *Psychological Reports, 100*(3, Pt 1), 894–896. https://doi-org.proxy-ub.rug.nl/10.2466/PR0.100.3.894-896

Meaux, L. T., Cox, J., & Kopkin, M. R. (2018). Saving damsels, sentencing deviants and selective chivalry decisions: Juror decision-making in an ambiguous assault case. *Psychiatry, Psychology and Law, 25*(5), 724–736. https://doi-org.proxy-ub.rug.nl/10.1080/13218719.2018.1474817

Ogungbamila, B., & Udegbe, I. B. (2014). Gender differences in the effects of perception of organizational injustice on workplace reactivity. *Europe's Journal of Psychology, 10*(1), 150-167.

Parker K., Funk C. (2017, December 14). Gender discrimination comes in many forms for today's working women. *Pew Research Center.* https://www.pewresearch.org/fact-tank/2017/12/14/gender-discrimination-comes-in-many-forms-for-todays-working-women/

Porter S., ten Brinke L. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology, 14*(1), 119–134. https://doi.org/10.1348/135532508X281520

Pozzulo, J. D., Dempsey, J., Maeder, E., & Allen, L. (2010). The Effects of Victim

Gender, Defendant Gender, and Defendant Age on Juror Decision Making. *Criminal Justice and Behavior, 37*(1), 47–63. https://doi.org/10.1177/0093854809344173

Reynolds, T., Howard, C., Sjåstad, H., Zhu, L., Okimoto, T. G., Baumeister, R. F., … Kim, J. (2020). Man up and take it: Gender bias in moral typecasting. *Organizational Behavior and Human Decision Processes, 161*, 120–141. doi:10.1016/j.obhdp.2020.05.002

Richeson, J. A., & Ambady, N. (2001). Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles*, *44*, 493-512.

Rotundo, M., Nguyen, D. H., & Sackett, P. R. (2001). A meta-analytic review of gender differences in perceptions of sexual harassment. *Journal of applied psychology, 86*(5), 914.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of personality and social psychology*, *87*(4), 494.G

Scheepers, D., Spears, R., Doosje, B., & Manstead, A. S. R. (2006). The social functions of ingroup bias: Creating, confirming, or changing social reality. *European Review of Social Psychology, 17*(1), 359–396. doi:10.1080/10463280601088773

Stockdale, M. S., Bisom-Rapp, S., O'Connor, M., & Gutek, B. A. (2004). Coming to terms with zero tolerance sexual harassment policies. *Journal of Forensic Psychology Practice, 4*(1), 65-78.

Voogt, A., & Klettke, B. (2017). The Effect of Gender on Perceptions of Credibility in Child Sexual Assault Cases: A Systematic Review. *Journal of Child Sexual Abuse, 26*(2), 195–212. doi:10.1080/10538712.2017.1280576

Waldo, C. R., Berdahl, J. L., & Fitzgerald, L. F. (1998). Are men sexually harassed? If so, by whom? *Law and Human Behavior, 22*(1), 59–79. https://doi.org/10.1023/A:1025776705629

Appendix A

Female Victim Vignette

Please read about Joan's complaint against Keith. Joan noted that the company treated her unfairly because of her gender. Joan notes that she has experienced and witnessed extensive sexist behavior in this place. She called them out for perpetuating a climate of sexism and hostilities toward women. For instance, she noted that as a woman, she was denied promotion, was forced to put up with managers' and others' bad jokes, and felt excluded for her gender. She specifically called out her colleague Keith at the company's anniversary party, who made a tasteless joke, but despite a large crowd, nobody heard it. Keith claims it was just an innocent joke and it was not meant to hurt anyone. HR has investigated the complaint but could not find clear evidence to substantiate Joan's claim and take action against Keith. HR is aware of only one other complaint, but that one directly contradicts Joan's claim and alleges that there is sexism against men (Sam's complaint against Marla).

Male Victim Vignette

Please read about Sam's complaint. Sam noted that the company treated him unfairly because of his gender. Sam notes that he has experienced and witnessed extensive sexist behavior in this place. He called them out for perpetuating a climate of sexism and hostilities toward men. For instance, he noted that as a man, he was denied promotion, was forced to put up with managers' and others' bad jokes, and felt excluded for his gender. He specifically called out his colleague Marla at the company's anniversary party, who made a tasteless joke, but despite a large crowd, nobody heard it. Marla claims it was just an innocent joke and it was not meant to hurt anyone. HR has investigated the complaint but could not find clear evidence to substantiate Sam's claim and take action against Marla. Moreover, HR is aware of only one other complaint, but that one

directly contradicts Sam's claim and alleges that there is sexism against women (Joan’s complaint against Keith).

Appendix B

Table 1

Correlations Between the Variables

		FV Severity	MV Severity	FV Discipline	MV Discipline	FV Dismissal	MV Dismissal
FV Severity	*r*	--					
	N	195					
MV Severity	*r*	.678**	--				
	p	<.001					
	N	194	194				
FV Discipline	*r*	.474**	.353**	--			
	p	<.001	<.001				
	N	179	178	180			
MV Discipline	*r*	.372**	.504**	.829**	--		
	p	<.001	<.001	<.001			
	N	179	178	171	179		
FV Dismissal	*r*	.364**	.234**	.489**	.380**	--	
	p	<.001	0.001	<.001	<.001		
	N	194	193	180	179	195	
MV Dismissal	*r*	.345**	.376**	.456**	.455**	.866**	--
	p	<.001	<.001	<.001	<.001	<.001	
	N	194	193	180	179	195	195

Note. FV denotes Female Victim. MV denotes Male Victim. r denotes Pearson's Correlation. p denotes sig. (2-tailed). ** Correlation is significant at the 0.001 level (2-tailed).

Figure 1

Normality Assumption Female Victim Severity

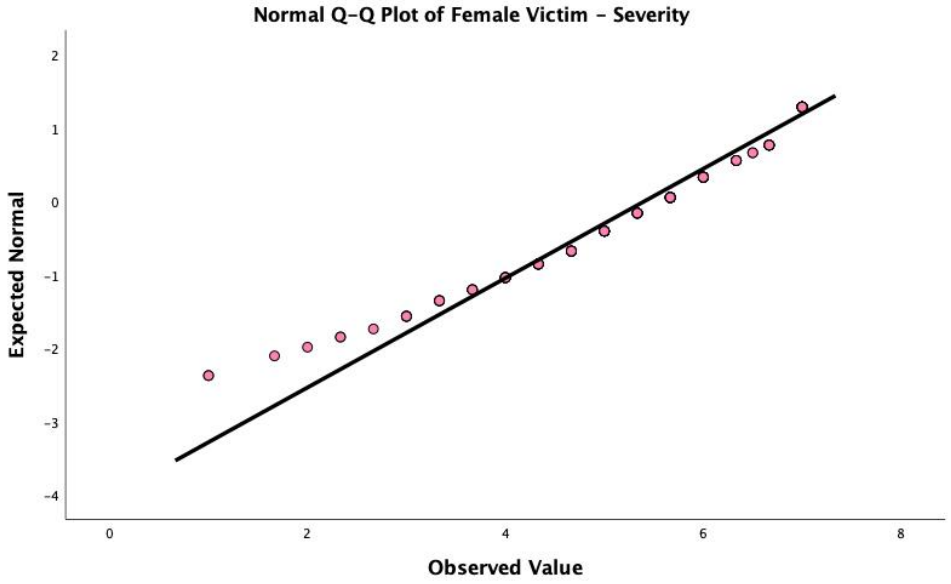

Figure 2

Normality Assumption Male Victim Severity

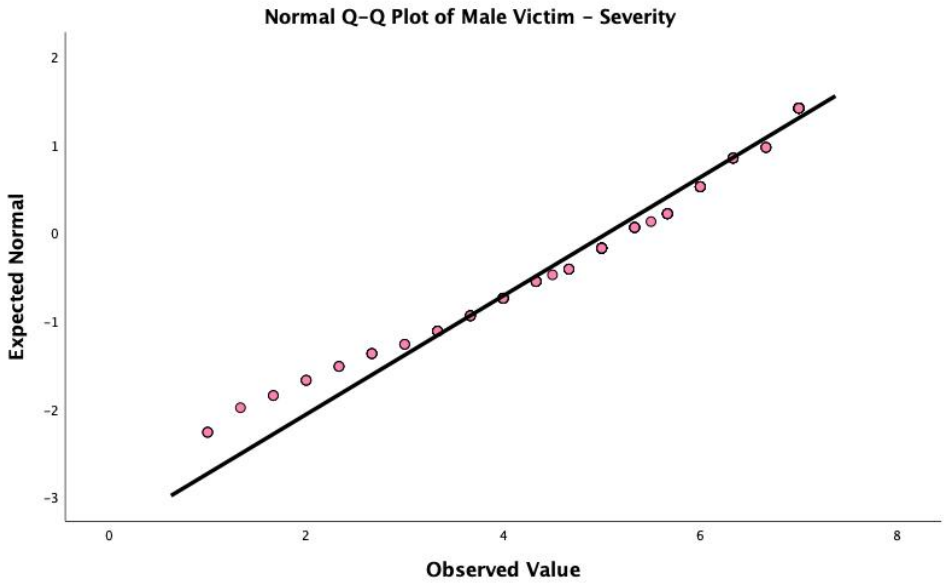

Figure 3

Normality Assumption Female Victim Discipline

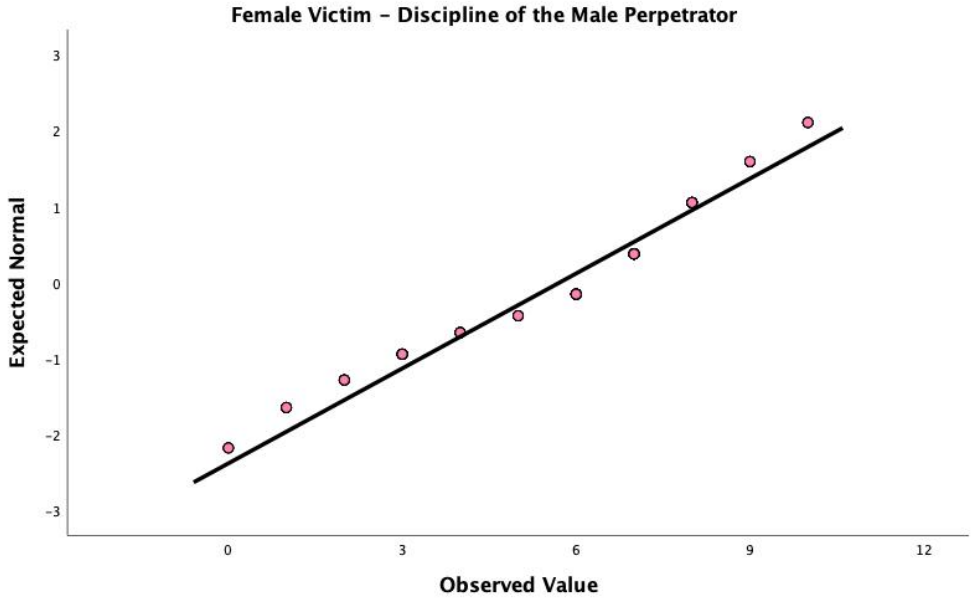

Figure 4

Normality Assumption Male Victim Discipline