# Registered Report: See me, judge me, pay me: Gendered effort moralization in work and care.

Roth<sup>1,2,3</sup>, L. H. O., Tissot<sup>1,4</sup>, T. T., Fischer<sup>3</sup>, T., Masak<sup>3</sup>, S., C.

<sup>3</sup>Department of Occupational, Economic, and Social Psychology, Faculty of Psychology,

University of Vienna, Austria

<sup>4</sup>Ghent University, Faculty of Psychology and Educational Sciences, Department of

Developmental, Personality, and Social Psychology, Ghent, Belgium

<sup>1</sup>the first two authors share first authorship

<sup>2</sup>corresponding author

## Authors note

Leopold Helmut Otto Roth<sup>1,2</sup>, 0000-0002-1120-4733 Tassilo Tom Tissot<sup>1</sup>, 0000-0002-6215-975X Thea Fischer, 0009-0002-0490-4894 Sophie Charlotte Masak, 0009-0005-1555-9258

Data, materials, code, and further associated documents: https://osf.io/s8ec5/

## Author contribution

1

Conceptualization: LR, TT, TF Data curation: LR, TT Formal analysis: LR Funding acquisition: LR, TF Investigation: LR, TT, TF, CM Methodology: LR, TT Project administration: LR, TT, TF, CM Visualization: LR, CM Writing - original draft: LR, TT, TF, CM Writing - review & editing: LR, TT, TF, CM

#### Abstract

The display of high effort at work is commonly rewarded with more positive moral judgements and increased cooperation partner attractiveness. This effect was shown to hold, even if higher effort is unrelated to better performance. Yet, current evidence is predominantly based on male agents, situated in the work context. This leaves a meaningful gap as it prohibits generalization to the full population and neglects critical aspects of our lives besides work, such as the care context (e.g., childcare or care for elders). To enhance our understanding of effort moralization and the impact of gender stereotypes on social judgment, we conducted two studies ( $N_1 = xxx$ ,  $N_2 = xxx$ ) testing the effect between genders in two contexts—work and care. [completion at Stage 2].

Keywords: Registered Report, effort moralization, gender, work, care, bias, judgment

					[				
Question	Hypothesis	Sampling plan	Analysis plan	Rationale for deciding the sensitivity	Interpretation given different outcomes	Theory that Formatierte Tabelle be shown wrong by the outcomes			
			Aim 1: Replication	of core effect					
Can we replicate the effort moralization effect?	Individuals who invest <b>higher</b> <b>effort</b> in their <b>work</b> are judged <b>higher in</b> <b>morality.</b> Individuals who invest <b>higher</b> <b>effort</b> in their <b>care</b> <b>work</b> are judged <b>higher in</b> <b>morality.</b>	We will collect data through <b>Prolific</b> . The <b>total required</b> sample size is N = 648, which we will <b>oversample</b> to N = 700. The required sample size per t-test is $N = 272$ .	Using a <b>one-sided dependent</b> Welch's t-test and respective Bayes Factor, we will test for differences in perceived moral character (core goodness and value commitment). We will further test for differences in perceived warmth, perceived competence <sub>1</sub> and pay deservingness. Yet, prior research highlighted variance in these more distal measures.	Based on the <b>smallest effect</b> <b>size of interest</b> (cite Lakens, 2022) approach, we aim to power for a small effect $d =$ <b>0.20</b> (Cohen, 1988) ( $a = .05$ , $1-\beta = .95$ , one-tailed). This was computed, using G*Power 3.1.9.7 [see supplemental material, <u>https://osf.io/s8ec5/</u> ].	If the effect is not found, the effort moralization effect is <b>not replicated</b> in the target magnitude. This can be due to the absence of the effect or due to the pooling of genders, which is tested in the following steps. If the effect is not found, it is potentially <b>not</b> <b>generalizable</b> to care work. Yet, the following analyses test the results in a more fine-grained	Effort moralization theory's generalizability could be shown undetectable under the current conditions of the study. Effort moralization is potentially not generalizable to the care context.			
	manner.								
	Formatierte Tabelle								
Are there differences in effort	Moral character judgment differs by gender and	In study 1 (work context), we	Using <b>mixed-ANOVA</b> with 2 (gender: female/male) x 2 (effort: high/low) design.	Based on the <b>smallest effect</b> <b>size of interest</b> (Lakens, 2022) approach, we aim to power for	The ANOVA- <del>and post- hoc tests</del> can <b>illustrate</b> whether gender and/or	If effort does Verbundene Zellen affect moral character jude Verbundene Zellen			

# Design Table

3

moralization in the <b>work</b> <b>context</b> by gender and effort?	effort. One of the hypothesized models (female high morality, male repercussions, female repercussions) explains the data best	will sample N = <b>350</b> individuals (computed N = <b>324</b> )	Gender serves as a between- subject factor, and effort is a within-subject factor. Post- hoe tests are computed, using Tuckey HSD correctionWe further test the interaction of both terms. The respective Bayes Factor of the ANOVA model is computed against the null- modelfor each term.	a small effect $\eta^2 = .01$ (Cohen, 1988) ( $a = .05, 1-\beta = .95,$ 2(gender)x2(effort)). This was computed, using G*Power 3.1.9.7 [see supplemental material, https://osf.io/s8ec5/].	effort differentially influence moral character judgment in the work context. The model comparison will inform us, whether one of the hypothesized models optimally explains the data or whether a different pattern was the data- generating mechanism.	the effect is potentially not replicable in this context. If it shows differences by gender, the effect is potentially hat forr heterogeneou Staaten) between genders (female/male). A minimum of two hypothesized models will be sub- optimal for explaining the collected data.	<b>natiert:</b> Schriftart: Fett, Englisch (Vereinigte
Are there differences in effort moralization in the <b>care</b> <b>context</b> by gender and effort?	Moral character judgment differs by gender and effort.	In study 2 (care context), we will sample $N =$ <b>350</b> individuals (computed $N =$ <b>324</b> )	Using mixed-ANOVA with 2 (gender: female/male) x 2 (effort: high/low) design. Gender serves as a between- subject factor, and effort is a within-subject factor. Post- hoe tests are computed, using Tukey HSD correction and We further test the interaction of both terms.	Based on the smallest effect size of interest (Lakens, 2022) approach, we aim to power for a small effect $\eta^2 = .01$ (cite Cohen, 1988) ( $a = .05, 1-\beta =$ .95, 2(gender)x2(effort)). This was computed, using G*Power 3.1.9.7 [see supplemental material, https://osf.io/s8ec5/].	The ANOVA and post- hoc tests can <b>illustrate</b> <b>whether gender and/or</b> <b>effort</b> differentially influence <b>moral</b> <b>character</b> judgment in the <b>care context</b> .	If effort doesn't affect moral character judgment, the effect is potentially not replicable in this context. If it shows differences by gender, the effect is potentially	

	One of the hypothesized model (care model) explains the data best		The respective <b>Bayes Factor</b> is computed for each term.		The model comparison will inform us, whether the hypothesized model optimally explains the data or whether a different pattern was the data generating mechanism.	heterogeneous between genc hat for (female/male (Vereini The hypothesized model is potentially suboptimal for explaining the data.	matiert: Schriftart: Nicht Fett, Englisch gte Staaten)
		Aim 3: co	ooperation partner satisfaction	as a function of gender and effo	ort	Format	ierte Tabelle
Do gender and effort influence cooperation satisfaction?	work context: cooperation satisfaction is predicted by gender and effort	In each study, we will sample sample N = 350 individuals (computed $N = 272$ )	Using <b>mixed-ANOVA</b> with 2 (gender: female/male) x 2 (effort: high/low) design. <b>Gender</b> serves as a <b>between-</b> <b>subject</b> factor, and <b>effort</b> is a <b>within-subject</b> factor. <b>Post-</b> <b>hoe</b> tests are We further test the interaction of both terms. <u>The respective <b>Bayes Factor</b></u> is computed, using <b>Tukey</b> <b>HSD correction</b> for each term.	Based on the <b>smallest effect</b> <b>size of interest</b> (Lakens, 2022) approach, we aim to power for a small effect $\eta^2 = .01$ (Cohen, 1988) ( $a = .05$ , $1-\beta = .95$ , 2(gender)x2(effort)). This was computed, using G*Power 3.1.9.7 [see supplemental material, <u>https://osf.io/s8ec5/</u> ].	We will be informed as to which degree <b>effort</b> is crucial for cooperation satisfaction <b>between</b> <b>women and men</b> in the <b>work context</b> .	Effort might not be a meaningful predictor of cooperation satisfaction. Further, there might not be differences between females and males.	
	care context: cooperation		Using <b>mixed-ANOVA</b> with 2 (gender: female/male) x 2		We will be informed as to which degree <b>effort</b> is		

	satisfaction is predicted by gender and effort		(effort: high/low) design. Gender serves as a between- subject factor, and effort is a within-subject factor. Post- hoe tests are We further test the interaction of both terms. The respective Bayes Factor is computed, using Tukey HSD correction for each term.		crucial for cooperation satisfaction <b>between</b> <b>women and men</b> in the <b>care context</b> .	
	<u>Explor</u>	atory Analysis: a	are differences in effort moraliz	zation moderated by gender nor	r <u>m endorsement</u>	
Are differences in effort moralization between genders moderated by gender norm endorsement?	This will be tested in the work and care context	This exploratory analysis will be performed on the computed samples size of Aim 2	Using multilevel modelling, we will test the effect of the interaction of gender and gender norm endorsement and the main effect of effort on moral judgement (core goodness & value commitment). Imer-formula: morality ~ gender*gender_norm + effort + (1 subject) We will further compare the Bayes Factor of the model	This exploratory analysis will be performed on the computed samples size of <b>Aim 2</b>	We will be informed whether gender norm endorsement moderates the influence of gender on effort moralization and whether the effect is generalizable on work and care contexts.	The effect is either generalizable on both contexts, context dependent or not observable with the present data.

1				
		against the model without		
		gender norm endorsement.		
		gender norm endorsement.		

#### Introduction

#### The effort moralization effect

Social judgment is crucial in daily life. People frequently encounter strangers and have to make quick inferences about their character, such as deciding whether it is safe to sit next to someone on the bus. Considering how important these decisions are, it is notable that we need to rely on rough, incomplete information to make such critical assessments - it wouldn't be feasible to administer a personality test to every passenger on the bus before choosing where to sit. We navigate such social interactions as cognitive misers, using simple processing mechanisms to reduce cognitive load (Fiske & Taylor, 1991). Instead of seeking complete information, we rely on environmental cues (e.g. valence of facial expressions; Fox et al., 2002), stereotypes (Aronson et al., 2021), heuristics and resulting cognitive biases (Tversky & Kahneman, 1974), and personal learning experiences (Behrens et al., 2008).

One factor that plays a dominant role in the perception of other people is moral information (Wojciszke, 2005; Goodwin et al., 2014; Brambilla & Leach, 2014). In this context, a particular bias has gained recent attention: the effort moralization effect (Fwu et al., 2014; Bigman & Tamir, 2016; Amos et al., 2019; Celniker et al., 2023). It describes the tendency of observers to make moral character judgments based on the observed effort a person puts in a given behavior. The perceived intensity of effort amplifies moral judgments: actions perceived as "good" appear even more virtuous, while "bad" behaviors seem worse the more effort is involved (Bigman & Tamir, 2016). For example, it has been shown that donations of time are perceived as a greater emotional investment, and therefore better moral character, compared to donations of money (Reed et al., 2007; Johnson & Park, 2021).

Interestingly, the effort moralization effect persists even when the added effort does not lead to increased performance (e.g., better outcomes at work) (Celniker et al., 2023). This points to the interpretation that the exertion of effort is valued by itself, rather than its

practical benefits. This observation was replicated well, yet it appears to vary between cultures in magnitude (France: d = .38, US: d = .60, South Korea: d = .71, add Mexiko/Germany) (Celniker et al., 2023; Tissot & Roth, 2024).

Further, it was shown that the display of high effort - contrasted with low effort for the same outcome - led to an increased chance of being selected as a cooperation partner in a follow-up trust game (Celniker et al., 2023), which has meaningful implications, especially for the work and education environmentscareer context.

#### Current blind spotsgaps in the effort moralization literature: context and gender

Prior literature has mostly focused on two types of contexts, in which effort moralization comes to play: work contexts (Amos et al., 2019; Celniker et al., 2023) and charity or helping behavior (Bigman & Tamir, 2016; Celniker et al., 2023). These contexts are justified targets, as these are impactful domains in our lives and commonly demand effort. Yet, it left the large domain of unpaid care work uncovered, which is estimated to make up 245 hours of annual work for the average American citizen (Mason & Robbins, 2024). Twothirds of care work (65%) is done by women (Mason & Robbins, 2024), and often goes with little societal recognition (Antonopoulos, 2008) and high mental load (Dean et al., 2021), while it surpasses the value of \$1 trillion dollar in the US per year (National Partnership for Women & Families, 2024). Further - to our knowledge - the literature on the effort moralization effect focused on either male or gender-neutral (e.g., Person A) vignettes and excluded female agents from described scenarios. Hence, investigating the role of an additional critical context, as well as between-gender effects and differential effects on moral character judgment, appears warranted for the generalizability of the effect. Understanding gender bias in the effort moralization effect is crucial for addressing inequalities (e.g., reinforcement of traditional gender roles).

Celniker et al. (2023) discovered that individuals who exert more effort to achieve the same performance in widget-making are more likely to be chosen as partners in a trust game. However, such freedom of partner choice is often absent in real-world situationscooperation partners are not always freely chosen but can be assigned as well (e.g-, project assignments in the workplace). We address this gap in We, therefore, extend the literature by assessing individuals' satisfaction with assigned, rather than<u>instead of</u> freely chosen, partners. This provides <u>additional</u> insights that more accurately reflect the cooperative dynamics frequently found in everyday life,

Suggested theoretical models of <u>Gendered stereotyping in moral character</u>judgment and <u>effort perceptions</u>

We developed four different models <u>derived from As shown in prior research</u>, <u>social judgment is not immune</u> to depict possible patterns of effort moralization for men and women in the work and care context. While the Female High Morality Model (A) emphasizes the heightened moral-influences of stereotyping, including gender biases. These extend to <u>differing expectations placedof behavior and personality based</u> on women, the Male Repercussion Model (B) and Female Repercussion Model (C) focus on the backlash men and women face for not meeting traditional standards. The Care Model (D) conceptualizes these dynamics within the care work context. Feldfunktion geändert

hat formatiert: Englisch (Vereinigte Staaten), Hervorheben

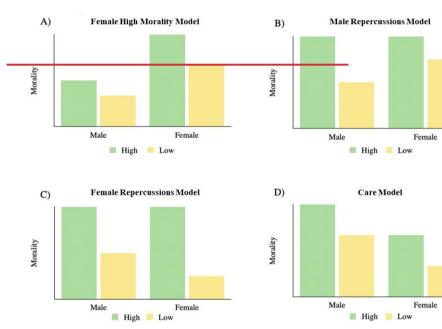
hat formatiert: Schriftart: Times New Roman, 12 Pt., Fett, Englisch (Vereinigte Staaten)

hat formatiert: Schriftart: Times New Roman, 12 Pt., Fett, Englisch (Vereinigte Staaten)

Formatiert: Standard

hat formatiert: Schriftart: Times New Roman, 12 Pt., Fett, Englisch (Vereinigte Staaten)

#### Figure 1



#### Suggested Models of Moral Character Judgment

Models of character judgment in the work context

#### Female high morality model (A)

Introduced in the "Big Two of Personality" framework (Bakan, 1966), communion and agency have been shown to encompass the majority of gender stereotypes (Haines et al., 2016). While communion encompasses qualities such as being caring, helpful, and sympathetic, agency refers to traits like assertiveness, self reliance, and independence (Abele et al., 2008). Commonlya person's gender. For instance, while men are often seen as more

agentic, women are perceived as more communal and men as more agentic (Hentschel et al., 2019). The effort shown at work might be perceived as a manifestation of these stereotype dimensions. For the agentic man, effort is a means of self-advancement and assertion. For the communal woman, it is interpreted(e.g., caring or helpful) (Hentschel et al., 2019). These expectations may inform differences in effort moralization, examplewise through the lens of striving for community, cooperation, and connectionbacklashing, and can differ between contexts.

Since morality is perceived as a female core trait and therefore a standard for what to evaluate them on, effort might be moralized stronger for female agents, i.e. they will be judged more positively in terms of morality if they display high effort behavior. However, behavior inconsistent with stereotypes can lead to backlash effects such as social and economic sanctions, i.e. women — who are expected to be moral — are punished more severely for ethical violations in the workplace (Kennedy et al., 2016; Rudman, 1998). In conclusion, the female high morality model posits that women are more strongly rewarded for high effort and more severely punished for low effort.

#### Male repercussion model (B)

An alternative viewpoint is that men and women might be judged similarly for demonstrating high effort at work. However, traditional gender norms situate strenuous work and career dedication as fundamental aspects of the male identity. Consequently, if men fail to display the expected agentic stereotype they may face repercussions for contradicting these expectations (by displaying low effort). For example, research indicates that men who exhibit modest behavior are perceived as less favorable than women who display identical behaviors when applying for a managerial position (Moss-Racusin et al., 2010). In summary, this model postulates that both genders are rewarded for high levels of effort, yet the consequences for

low effort differ. Men are subjected to more severe repercussions for low effort in the workplace.

#### Female repercussion model (C)

It can be similarly presumed that, within a context wherein men and women are evaluated equally for their high level of effort at work, women are subjected to greater scrutiny for their lack of effort. As previously discussed in the female high morality model (A), women are expected to exemplify moral values but<u>This describes how expectations, for</u> instance, those formed by gender, can lead to differing social judgments (Rudman, 1998). Individuals who deviate from stereotypical behavior tend to be subjected to harsher sanctions. For example, women receive more severe disciplinary sanctions for ethical violations in the workplace (Kennedy et al., 2016; Rudman, 1998), whereas men face greater criticism for non-agentic behavior in leadership contexts (Moss-Racusin et al., 2010).

For effort moralization, these prior findings hold potential for differences in judgment between gender, effort levels, and social context. Male stereotypes of agentic behavior could cause stronger differences in moral judgment at work, as men are expected to work hard and autonomously.

<u>The interplay between effort moralization and gender</u> may be subjected to greater repercussions for contradicting the female stereotype. Following the female repercussion model, it can be predicted that men and women are evaluated similarly for their high level of effort at work, while women are subjected to more severe penalties for exhibiting low effort. *Model of character judgment in the care context* 

#### Care model (D)

The Care model offers a theoretical framework for understanding how traditional gender roles may shape perceptions of effort, particularly within the context of care work. Stereotypically, men are seen as breadwinners engaging in paid labor, while women are

associated with unpaid care work.expand to the caregiving contexts. Although gender roles have shifted over time , with more women participating inentering the U.S. workforce (Toossi & Morisi, 2017; Statista Research Department, 2024)(Toossi & Morisi, 2017) and men contributing strongermore to family labor (Sayer, 2016) women continue to bear a larger share of caregiving and household work (Charmes, 2019). We propose that these(Sayer, 2016), women do most of the care work (Charmes, 2019). These persistent stereotypes have implications for expectations shape how effort in the care context is moralized. For women, high effort in caregiving might not warrant special moral praise, as it is expected and aligns with their traditional role. For men, however, high effort in the care domain might be perceived as going above and beyond the traditional responsibilities associated with their role. In addition, a woman's low effort in care work could be perceived as neglect of her responsibilities, leading to repercussions. In contrast, men's low effort in care work might be generally expected and therefore seen as neutral, neither drawing praise nor criticism. This is supported by researchefforts are perceived. Research on double standards for mothers and fathers: has shown that mothers face harsher criticism for "too little"low care work,effort while fathers receive greater praise for being involved (Deutsch & Saxon, 1998). (Deutsch & Saxon, 1998).

In sum, existing literature indicates that gender biases likely play a role in effort moralization. The specific goals are outlined below.

#### **Current study**

The current project focused on three core aims. I) To test the replicability of the effort moralization effect and explore its generalizability to the care work context. II) To examine how the effort moralization effect interacts with gender across different contexts and to

**hat formatiert:** Schriftart: Kursiv, Englisch (Vereinigte Staaten)

#### evaluate which of the proposed models explains the data patterns adequately.. III) To

investigate whether cooperation partner satisfaction differs by gender and effort.

#### Method

# Note that all data is simulated in the stage 1 manuscript and will be replaced with the real data in the stage 2 manuscript

#### Sample and sample size

Using the smallest effect size of interest approach (Lakens, 2022), we powered both studies to detect a small effect (Cohen, 1988) in <u>ma 2X2 mixed</u> ANOVA-with four groups<sup>4</sup> ( $\eta^2 = .01$ ,  $\alpha = .05$ ,  $1-\beta = .95$ ). This resulted in a minimum sample size of N = 324 per study, which we decided to oversample to N = 350 per study. We conducted a second power analysis for the interaction effect with the same parameters (d = 0.20), resulting in a similar required sample size (N = 325) to countercheck between computation tools. Hence, the total target sample size was N = 700. The computation was done using G\*Power 3.1.9.7 (Faul et al., 2009) and IntXPower (Sommet et al., 2023), documented in the supplemental material (<u>https://osf.io/s8ec5/}.)<sup>2</sup></u>. Participants were recruited via Prolific and consisted of US participants.

#### Table 1

#### Descriptives by Study

	Ν	age: M (SD)	range	%female <	Formatierte Tabelle
overall	321	68.14 (30.96)	18-120	25.50	-
Study 1 (work)	164	66.25 (29.78)	18-120	31.71	-
Study 2 (care)	157	70.12 (32.12)	18-120	19.11	

<sup>&</sup>lt;sup>4</sup> for data analysis, we use an 2X2 mixed effect ANOVA type III, instead of an four group between subject ANOVA. Given the number of potential outcome patterns (see Figure 1), we did not feel comfortable to make a elaim on the specific form of mean patterns for the power analysis. Yet, powering for a small effect appeared sufficiently conservative, given the uncertainties, while not basing it on false claims of certainty.
<sup>2</sup> We are aware that some patterns of interaction terms potentially require larger samples. Given that the pattern

<sup>&</sup>lt;sup>2</sup> We are aware that some patterns of interaction terms potentially require larger samples. Given that the pattern is not known at this point, it can happen that some interaction forms might not be sufficiently powered through our sample.

#### Materials

To build on previous research, we employed similar instruments and adapted prior vignettes to fit the current studies (Celniker et al., 2023). Specifically, we designed a new vignette tailored to the caregiving context and adapted the work vignette to feature less stereotypical tasks, aiming to minimize potential distortions through biases associated with traditional male or female roles (office instead of factory scenario). This was particularly important as we intended to include both male and female agents in the scenarios.

Building on previous research, we assessed the perceived morality of agents using 13 trait items (Celniker et al., 2023) that have been demonstrated to distinguish between two types of moral virtues (Piazza et al., 2014). While core goodness traits like kindness are universally good, the moral valence of value commitment traits like dedication depends on the context – a kind murderer is "better" than an unkind one, butwhile a dedicated murderer is "worse" than an undedicated one. All trait items were rated on a 7-point scale.

Following the procedure of Celniker et al. (2023), warmth and competence, two universal dimensions of social cognition for anticipating interdependence and status, were assessed with one item each on a 7-point scale (Fiske et al., 2007).

The perceived effort, quality, difficulty, and work value were measured with single items rated on a 7-point scale. They served as manipulation checks.

The item assessing the pay deservingness of each agent differed between the work and care context study. In the work context study, participants responded on a sliding scale, anchored at a midpoint that reflected a realistic average office worker salary in the US. In the care context study, no reference point was provided, given that this work is typically unpaid. Instead, participants could freely choose a salary between \$0 and \$50. This allowed us to assess the perceived value of care work. For estimating realistic salaries in the US we relied See me, judge me, pay me on data shared on the webpage of the ERI Economic Research Institute (https://www.erieri.com). Further, we assessed how satisfied participants would be on a 7-point scale to have either agent as an assigned cooperation partner in a work project (work context) or organizing a charity event (care context). In addition, for exploratory purposes, we incorporated a short version of the gender role belief scale into our study to explore potential moderating effects of traditional gender role endorsement on effort moralization (Brown & Gladstone, 2012). All materials are

available in the supplemental materials (https://osf.io/s8ec5/).

Formatiert: Einzug: Erste Zeile: 1,27 cm

## Table 2

#### **Overview** of Measures

overview of m	easures		
Construct ( <i>n</i> items)	Example item	Low anchor	High anchor • Form
core goodness (6) <sup>a</sup>	honest	does not describe X well	describes X extremely well
value commitment (7) <sup>a</sup>	dedicated	does not describe X well	describes X extremely well
competence/warmth (2) <sup>a</sup>	competent	does not describe X well	describes X extremely well
effort (1) <sup>b</sup>	How much effort do	no effort at all	a lot of effort

	you think X puts into his/her (care) work?		
quality (1) <sup>c</sup>	In your opinion, how well does X perform his/her (care) work?	very bad	very good
difficulty (1) <sup>c</sup>	Compared to other jobs/ care work, how difficult is X's (care) work.	not at all difficult	extremely difficult
work value (1) <sup>c</sup>	How valuable do you think X's (care) work is?	not valuable at all	extremely valuable
pay deservingness (work) (1) <sup>a</sup>	The average office worker at the company makes \$24 an hour. How much do you think X should make per hour?	\$12	\$36
pay deservingness (care) (1) <sup>a</sup>	Imagine that X was paid for his/her care work. How much should s/he be paid per hour?	\$0	\$50
collaboration partner choice (1) <sup>a</sup>	[] Please indicate how satisfied you would be to work with either X or Y?	extremely dissatisfied	extremely satisfied
<u>gender role beliefs</u>	Women with children should not work outside the home if they don't have to financially.	Strongly disagree	Strongly agree

Note. <sup>a</sup>These variables are the focal dependent measures, <sup>b</sup>This measure serves as

manipulation check and exclusion criterion, "These measures serve as manipulation check but

not as exclusion criterion.

#### Procedure

We conducted two independent studies to compare responses to a work vignette and a care vignette. The data was collected in two separate samples at month/year [Stage 2], with participants from one study being excluded from participating in the other. In both studies, after providing informed consent, participants were presented with a scenario from either the work or care context, depending on the study.

The vignettes featured two individuals – either male or female – who perform the exact same tasks at the same quality level but differ in the amount of effort required. For example, the work context vignette reads as follows<sup>3</sup>:

Anna and Sophie work at the same company and process similar orders in the company's office. Both Anna and Sophie are able to process approximately three orders per hour, which means they complete one case every 20 minutes. The average value of a completed case for the company is \$50.00. Quality control inspections indicate that 96% of Anna's and Sophie's orders are error-free and complete. On average, Anna and Sophie each process correct orders worth \$144 per hour.

For Anna, processing orders requires minimal effort — although she works as quickly as possible, she finds the work easy.

For Sophie, however, processing orders requires a lot of effort — although she works as quickly as possible, she finds the work hard.

After reading the vignette, participants completed a series of dependent measures for each featured individual in randomized order. <u>Subsequently, participants responded to the items constituting the short version of the gender role belief scale (Brown & Gladstone, 2012).</u> Within each study, gender (male vs. female names in the vignette) served as the between-subject factor, and effort (high vs. low) was the within-subject factor.

Both studies took approximately 7 minutes per participant, and data were collected via Prolific. Participants received compensation according to the platform's standard rates [in Stage 2].

<sup>&</sup>lt;sup>3</sup> vignettes were designed to reduce stereotyped associations. Hence we adapted the vignette by Celniker et al. (2023) from a factory to an office setting and designed the care vignette in a way that non-relational tasks are in the foreground (e.g., lawn mowing instead of emotional support).

#### Data cleaning

We applied several measures to ensure high data quality. To ensure valid responses, participants who self-reported insufficient English proficiency (below "very good") were excluded. Participants who failed one of two attention checks embedded within the study were excluded from the analysis. The probability of passing both attention checks by random guessing was  $(\frac{l}{7} \times \frac{l}{7} =) 2.04\%$ . Participants who completed the study 3 standard deviations (SD) faster than the average participant were excluded, as this can indicate insufficient attention to the task. There was no exclusion for slow participation. In line with the procedure by Celniker et al. (2023) and Tissot & Roth (2024) we further excluded participants who rated the low-effort condition as equally or more effortful than the high-effort condition. Participants who did not complete the study were excluded from the final analysis. The number of exclusions by reason and sample is documented in the supplemental material [added in Stage 2].

## Data analysis

#### Aim 1: Replication of core effect

To test whether the original effort moralization effect could be replicated in the work context and generalized to the care context, we conducted a series of dependent, one-sided Welch's t-tests, comparing moral judgments of the described agents between high- and loweffort conditions. We further computed the respective effect size (Cohen's *d*) and Bayes Factor (BF<sub>10</sub>). Additionally, we compared perceived warmth, competence, and pay deservingness between high and low effort agents.

#### Aim 2: Moral character as a function of gender, context, and effort

To examine the effects of gender and effort on moral character judgments—in both work and care contexts—we used a mixed effect ANOVA (between-subjects factor: gender, within-subjects factor: effort). Post hoe tests were conducted using Tukey's HSD correction,

along) with an interaction term. For all terms, the respective Bayes Factors (BF10)-) was

computed to quantify evidence of absence and presence of effects.

We compared the results against the proposed models in the respective context (work:

model 1-3, care: model 4).

#### WeTable 3

Patterns by Theoretical Model

Ð	model	order
4	female high morality	FhE > FIE > MhE > MIE
2	male repercussion	FhE = MhE > FIE > MIE
3	female repercussion	FhE = MhE > MIE > FIE
4	care	MhE > FhE = MIE > FIE

*Note.* M = male, F = female, hE = high effort, lE = low effort

Finally, we applied the same mixed-effects ANOVA procedure to participants'

suggested hourly payment to test for evidence of gender pay gaps.

## Aim 3: Cooperation partner satisfaction as a function of gender and effort

We used the same mixed-effects ANOVA procedure (as described in Aim 2) to

compare satisfaction with assigned cooperation partners.

#### Results

Note that all data is simulated in the stage 1 manuscript and will be replaced with the real data in the stage 2 manuscript

While not part of the Stage 1 manuscript, we illustrate first plans for how to present the data after collection

## Aim 1: Replication of core effect

We conducted a series of dependent, one-sided Welch's t-tests to test the replicability

of the core effort moralization effect in the work context and the generalizability of the effect

in the care context. [will be completed in Stage 2 manuscript].

#### Table 4

## Table 3

## Comparison of Moral Character Judgment by Effort and Context

	low effort: M (SD)	high effort: M (SD)	d [95% CI]	BF <sub>10</sub> Formatierte Tabelle
		work context		
core goodness	4.03 (0.82)	4.00 (0.76)	0.02 [-0.13, 0.17]	0.09
value commitment	3.98 (0.73)	3.99 (0.76)	-0.01 [-0.16, 0.14]	0.08
warmth	3.94 (2.07)	4.24 (1.96)	-0.10 [-0.25, 0.04]	0.21
competence	3.84 (2.03)	4.13 (1.89)	-0.11 [-0.26, 0.03]	0.24
pay deservingness	24.37 (7.09)	23.77 (7.07)	0.05 [-0.09, 0.21]	0.11
cooperation	4.09 (1.98)	3.93 (2.04)	0.05 [-0.09, 0.21]	0.11
		care context		
core goodness	4.08 (0.83)	3.95 (0.83)	0.10 [-0.04, 0.26]	0.26
value commitment	3.97 (0.73)	4.08 (0.75)	-0.09 [-0.25, 0.05]	0.05
warmth	4.22 (1.90)	3.91 (1.90)	0.11 [-0.04, 0.27]	0.27
competence	4.14 (1.98)	4.00 (1.91)	0.05 [-0.10, 0.20]	0.20
pay deservingness	25.85 (14.94)	26.62 (15.17)	-0.03 [-0.19, 0.12]	0.12
cooperation	4.15 (1.96)	4.03 (1.90)	0.04 [-0.11, 0.19]	0.19

*Note*. \*p < .05, \*\*p < .01, \*\*\*p < .001, Cohen's  $d \ge .20$  (smallest effect size of interest) in **bold print**.

## Aim 2: Moral character as a function of gender, context, and effort

#### work context

We fitted the 2x2 mixed ANOVA model (between: gender, within: effort) and

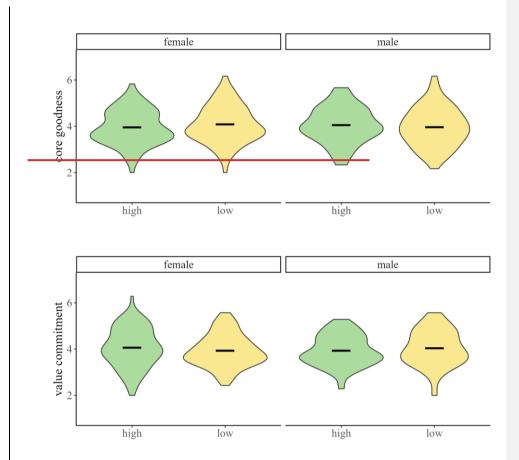
computed Tukey HSD post-hoc tests,), for which we further computed Bayes Factors.

## Figure 1

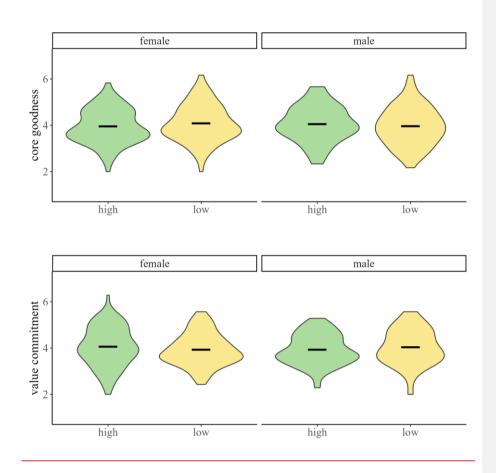
## Figure 2

Work Context: Mean Differences in Moral Judgment by Effort and Gender

See me, judge me, pay me



See me, judge me, pay me



care context

We fitted the 2x2 mixed ANOVA model (between: gender, within: effort) and

computed Tukey HSD post hoc tests,), for which we further computed Bayes Factors.

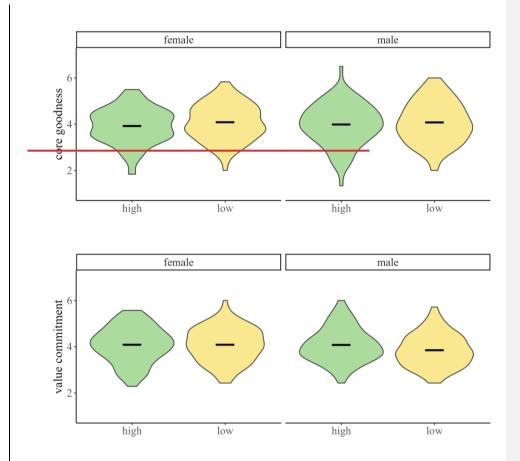
hat formatiert: Schriftart: Nicht Fett, Englisch (Vereinigte Staaten)

Formatiert: Einzug: Erste Zeile: 1,27 cm

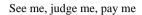
## Figure 2

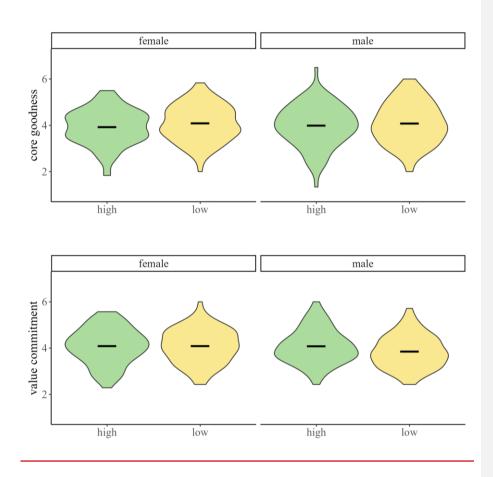
Figure 3

Care Context: Mean Differences in Moral Judgment by Effort and Gender



26





Further, we assessed suggested hourly salaries in both contexts to test for gender effects and whether effort differentially affected these between genders. [will be completed in Stage 2 manuscript].

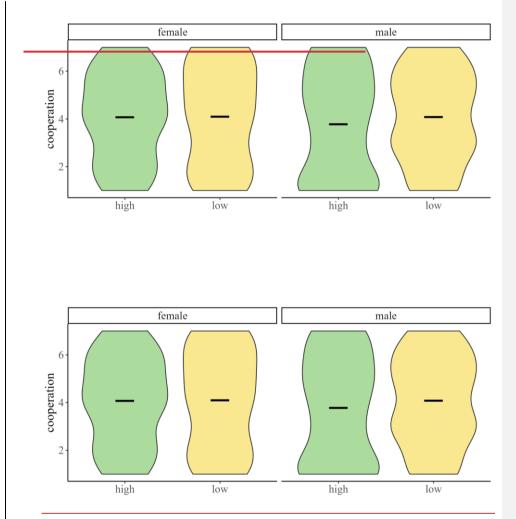
## Aim 3: Cooperation partner satisfaction as a function of gender and effort

To test whether perceived effort influences the satisfaction with assigned partners in cooperation scenarios, we used mixed-effect ANOVA to test the effect of effort (high effort/low effort) as a within-factor and gender (female/male) as a between-factor. The analysis was conducted by context (work/care). [will be completed in Stage 2 manuscript].

## Work context

# Figure 4<u>3</u>

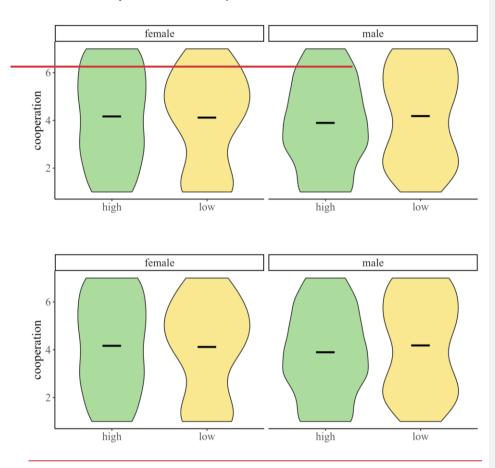
Work Context: Cooperation Partner Satisfaction



## Care context

Figure <mark>54</mark>





Care Context: Cooperation Partner Satisfaction

Exploratory Analysis: are differences in effort moralization moderated by gender norm

endorsement

Formatiert: Zeilenabstand: Doppelt, Keine Absatzkontrolle

#### Discussion

## Summary

#### Limitations

Conclusion

## **Competing interests**

The authors report no conflict of interest.

## Ethic approvement

The study was approved by the Departmental Review Board (DRB) of the Faculty of Psychology, Department of Occupational, Economic, and Social Psychology (2024/M/009).

## References

- Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, 38(7), 1202–1217. https://doi.org/10.1002/ejsp.575.
- Amos, C., Zhang, L., & Read, D. (2019). Hardworking as a Heuristic for Moral Character: Why We Attribute Moral Values to Those Who Work Hard and Its Implications. Journal of Business Ethics, 158(4), 1047–1062. https://doi.org/10.1007/s10551-017-3725-x
- Antonopoulos, R. (2008). The Unpaid Care Work-Paid Work Connection. <u>SSRN Electronic</u> Journal, https://doi.org/10.2139/ssrn.1176661
- Aronson, E., Wilson, T. D., Akert, R. M., & Sommers, S. R. (2021). *Social psychology* (Tenth edition. Global edition). Pearson Education Limited.
- Bakan, D. (1966). *Duality of human existence: An essay on psychology and religion*, Beacon Press.
- Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. <u>Nature</u> <u>456</u>(7219), 245–249. https://doi.org/10.1038/nature07538.
- Bigman, Y. E., & Tamir, M. (2016). The road to heaven is paved with effort: Perceived effort amplifies moral judgment. *Journal of Experimental Psychology: General*, 145(12), 1654–1669. https://doi.org/10.1037/xge0000230
- Brambilla, M., & Leach, C. W. (2014). On the Importance of Being Moral: The Distinctive Role of Morality in Social Judgment. *Social Cognition*, 32(4), 397–408. https://doi.org/10.1521/soco.2014.32.4.397
- Brown, M. J., & Gladstone, N. (2012). Development of a Short Version of the Gender Role Beliefs Scale. International Journal of Psychology and Behavioral Sciences, 2(5),

/	hat formatiert	
/	hat formatiert	
	hat formatiert	
$\parallel$	hat formatiert	
$\parallel$	hat formatiert	
//	hat formatiert	
//	hat formatiert	
1	hat formatiert	
1	hat formatiert	
$\parallel$	hat formatiert	
$\ $	hat formatiert	
$^{\prime\prime}$	hat formatiert	
7,	hat formatiert	
/	hat formatiert	
//	hat formatiert	
7	hat formatiert	
	hat formatiert	
7	hat formatiert	
/	hat formatiert	
	hat formatiert	
_	hat formatiert	<u></u>
/	hat formatiert	<u> </u>
/	hat formatiert	
	hat formatiert	
	hat formatiert	<u></u>
	hat formatiert	
<	hat formatiert	
_	hat formatiert	
	hat formatiert	<u> </u>
_	hat formatiert	<u> </u>
_	hat formatiert	<u> </u>
	hat formatiert	<u> </u>
/	hat formatiert	<u> </u>
/	hat formatiert	<u> </u>
		<u> </u>
	hat formatiert	<u> </u>
/		<u> </u>
1	hat formatiert	<u> </u>
1	hat formatiert	<u> </u>
	hat formatiert	<u></u>
	hat formatiert	[]

#### <u>154–158.</u>

- Celniker, J. B., Gregory, A., Koo, H. J., Piff, P. K., Ditto, P. H., & Shariff, A. F. (2023). The moralization of effort. *Journal of Experimental Psychology: General*, 152(1), 60–79. https://doi.org/10.1037/xge0001259
- Charmes, J. (2019). The unpaid care work and the labour market: An analysis of time use data based on the latest world compilation of time-use surveys. ILO.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed). L. Erlbaum Associates
- Dean, L., Churchill, B., & Ruppanner, L. (2021). The mental load: Building a deeper theoretical understanding of how cognitive and emotional labor over *load* women and mothers. *Community, Work & Family, 25*(1), 13–29. https://doi.org/10.1080/13668803.2021.2002813
- Deutsch, F. M., & Saxon, S. E. (1998). The Double Standard of Praise and Criticism for Mothers and Fathers. *Psychology of Women Quarterly*, 22(4), 665–683. https://doi.org/10.1111/j.1471-6402.1998.tb00184.x
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005
- Fiske, S. T., & Taylor, S. E. (1991). Social cognition (2. ed). McGraw-Hill,
- Fox, E., Russo, R., & Dutton, K. (2002). Attentional Bias for Threat: Evidence for Delayed Disengagement from Emotional Faces. *Cognition & Emotion*, 16(3), 355. https://doi.org/10.1080/02699930143000527
- Fwu, B., Wei, C.-F., Chen, S.-W., & Wang, H. (2014). Effort counts: The moral significance of effort in the patterns of credit assignment on math learning in the Confucian cultural context. *International Journal of Educational Development*, *39*, 157–162.

	hat formatiert	
	hat formatiert	
	hat formatiert	)
	hat formatiert	
	hat formatiert	<u> </u>
$\  \ $	hat formatiert	
	hat formatiert	
/ /	hat formatiert	
	hat formatiert	
$\langle \rangle$	hat formatiert	
	hat formatiert	)
	hat formatiert	
	hat formatiert	)
	hat formatiert	
	hat formatiert	$\neg$
$\sim$	hat formatiert	
	hat formatiert	
$\langle \rangle$	hat formatiert	
	hat formatiert	
$\backslash$	hat formatiert	<u> </u>
$\searrow$	hat formatiert	
$\sim$	hat formatiert	<u> </u>
$\mathbb{N}$		
$\langle N \rangle$	hat formatiert	
$\langle \rangle \rangle$	hat formatiert	
N	hat formatiert	
$\mathbb{N}$	hat formatiert	
	hat formatiert	
$\mathbb{N}/\mathbb{N}$	hat formatiert	
(   )	hat formatiert	
/////	hat formatiert	
11111	hat formatiert	
	hat formatiert	
	hat formatiert	
	hat formatiert	
/ <b>    </b>	hat formatiert	
	hat formatiert	
$\ $	hat formatiert	

32

https://doi.org/10.1016/j.ijedudev.2014.07.010

- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168. https://doi.org/10.1037/a0034726
- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing ... or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40(3), 353–363. https://doi.org/10.1177/0361684316634081
- Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The Multiple Dimensions of Gender Stereotypes: A Current Look at Men's and Women's Characterizations of Others and Themselves. *Frontiers in Psychology*, 10, 11.

https://doi.org/10.3389/fpsyg.2019.00011

- Johnson, S. G. B., & Park, S. Y. (2021). Moral signaling through donations of money and time. *Organizational Behavior and Human Decision Processes*, 165, 183–196. https://doi.org/10.1016/j.obhdp.2021.05.004
- Kennedy, J., McDonnell, M.-H., & Stephens, N. (2016). Does Gender Raise the Ethical Bar? Exploring the Punishment of Ethical Violations at Work. <u>Academy of Management</u> Proceedings, <u>2016</u>(1), <u>11664</u>. https://doi.org/10.5465/ambpp.2016.11664abstract
- Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267. https://doi.org/10.1525/collabra.33267
- Mason, J., & Robbins, K. G. (2024, June 27). Americans' Unpaid Caregiving is Worth More than \$1 Trillion Annually – and Women are Doing Two-Thirds of The Work, National Partnership for Women & Families. https://nationalpartnership.org/americans-unpaidcaregiving-worth-1-trillion-annually-women-two-thirds-work/
- Moss-Racusin, C., Phelan, J., & Rudman, L. (2010). When Men Break the Gender Rules: Status Incongruity and Backlash Against Modest Men. *Psychology of Men & Masculinity*, 11, 140–151. https://doi.org/10.1037/a0018093
- National Partnership for Women & Families. (2024). New Analysis Shows Unpaid Care Work in the U.S. is Worth More Than \$1 Trillion Each Year, National Partnership for

		_
A	hat formatiert	]
//	hat formatiert	]
///	hat formatiert	]
///	hat formatiert	
llλ	hat formatiert	)
///	hat formatiert	
Ζ,	hat formatiert	
/	hat formatiert	
/	hat formatiert	
//	hat formatiert	
// )	hat formatiert	
//	hat formatiert	
Ż	hat formatiert	Ä
Å	hat formatiert	
	hat formatiert	$\square$
	hat formatiert	
1/3	hat formatiert	
	hat formatiert	
1	hat formatiert	
	hat formatiert	
$\searrow$		<u> </u>
	hat formatiert	
	hat formatiert	
	hat formatiert	
	hat formatiert	<u> </u>
	hat formatiert	<u> </u>
$\backslash$	hat formatiert	<u> </u>
$\sum$	hat formatiert	
$\sum$	hat formatiert	]
	hat formatiert	]
$\mathbb{N}$	hat formatiert	]
	hat formatiert	]
	hat formatiert	)
\\\	hat formatiert	]
///	hat formatiert	]
	hat formatiert	
$\langle   l \rangle$	hat formatiert	<u> </u>
$\langle f \rangle$	hat formatiert	
, /]	hat formatiert	
N	hat formatiert	
$\sum$	hat formatiert	
	hat formatiert	
///	hat formatiert	
$\  \ $	hat formatiert	
///	hat formatiert	
	hat formatiert	
	hat formatiert	
Ì	hat formatiert	)

33

Women & Families. https://nationalpartnership.org/news\_post/new-analysisamericans-unpaid-care-work-worth-more-than-1-trillion-each-year/

- Piazza, J., Goodwin, G. P., Rozin, P., & Royzman, E. B. (2014). When a Virtue is Not a Virtue: Conditional Virtues in Moral Evaluation. *Social Cognition*, 32(6), 528–558. https://doi.org/10.1521/soco.2014.32.6.528
- Reed, A., Aquino, K., & Levy, E. (2007). Moral Identity and Judgments of Charitable Behaviors. *Journal of Marketing*, 71(1), 178–193.

https://doi.org/10.1509/jmkg.71.1.178

- Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology*, 74(3), 629–645. https://doi.org/10.1037/0022-3514.74.3.629
- Sayer, L. C. (2016). Trends in Women's and Men's Time Use, 1965–2012: Back to the Future? In S. M. McHale, V. King, J. Van Hook, & A. Booth (Eds.), *Gender and Couple Relationships* (Vol. 6, pp. 43–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-21635-5\_2
- Statista Research Department. (2024, May 7).- Sommet, N., Weissman, D. L., Cheutin, N., & Elliot, A. J. (2023). How Many Participants Do I Need to Test an Interaction?
   Conducting an Appropriate Power Analysis and Achieving Sufficient Power to Detect an Interaction. Advances in Methods and Practices in Psychological Science, 6(3), 25152459231178728. https://doi.org/10.1177/25152459231178728
- Statista Research Department. (2024, July 5). U.S. female labor force participation rate 1990-2023. Statista. https://www.statista.com/statistics/191737/us-female-laborforce-participation-rate-since-1990/
- Tissot, T. T., & Roth, L. H. O. (2024). *Is it Worth the Hustle? A Multi-Country Replication of the Effort Moralization Effect and an Extension to Generational Differences in the Appreciation of Effort*, https://osf.io/5tkfr

Toossi, M., & Morisi, T. L. (2017). BLS Spotlight on Statistics: Women in the Workforce

nat formatiert: Schriftart: Arial, 11 Pt.	
nat formatiert: Schriftart: Arial, 11 Pt.	
nat formatiert	(
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert: Schriftart: Arial, 11 Pt.	
nat formatiert	(
nat formatiert	(
nat formatiert	(
hat formatiert	(
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert	(
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert	(
hat formatiert	(
hat formatiert: Schriftart: Arial, 11 Pt.	

hat formatiert: Schriftart: Arial, 11 Pt.

hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert	
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert	
hat formatiert	
hat formatiert: Schriftart: Arial, 11 Pt.	
hat formatiert: Schriftart: Arial, 11 Pt.	
(,,,,,,,	

Before, During, and after the Great Recession, https://hdl.handle.net/1813/78334, Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Wojciszke, B. (2005). Morality and competence in person- and self-perception. *European Review of Social Psychology*, 16(1), 155–188.

https://doi.org/10.1080/10463280500229619

hat formatiert: Schriftart: Arial, 11 Pt. hat formatiert: Schriftart: Arial, 11 Pt.