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

Roth[1,2,3], L. H. O., Tissot[1,4], T. T., Fischer[3], T., Masak[3], S., C.

[3]Department of Occupational, Economic, and Social Psychology, Faculty of Psychology, University of Vienna, Austria

[4]~~Ghent University, Faculty of Psychology and Educational Sciences,~~ [4] Department of Developmental, Personality, and Social Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

[1]the first two authors share first authorship

[2]corresponding author

Authors note

Leopold Helmut Otto Roth[1,2], 0000-0002-1120-4733

Tassilo Tom Tissot[1], 0000-0002-6215-975X

Thea Fischer, 0009-0002-0490-4894

Sophie Charlotte Masak, 0009-0005-1555-9258

Author contribution

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

See me, judge me, pay me

Abstract

The display of high effort at work is commonly rewarded with more positive moral judgments 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

See me, judge me, pay me

See me, judge me, pay me

Design Table

Question	Hypothesis	Sampling plan	Analysis plan	Rationale for deciding the sensitivity	Interpretation given different outcomes	Theory that be shown wrong by the outcomes
Aim 1					
Aim 1: Replication of core effect						
Can we replicate the effort moralization effect?	Individuals who invest **higher effort** in their **work** are judged **higher in morality.**	We will collect data through **Prolific**. The **total required** sample size is *N* = **648**, which we will **oversample** to *N* = **700**. The required sample size per t-test is *N* = **272**.	Using a **one-sided dependent 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,** and **pay deservingness**. Yet, prior research highlighted variance in these more distal measures.	Based on the **smallest effect size of interest** ~~(cite Lakens, 2022)~~(Lakens, 2022) approach, we aim to power for a small effect *d* = **0.20** (Cohen, 1988) (*a* = .05, 1-*β* = .95, one-tailed). This was computed using G*Power 3.1.9.7 [see supplemental material, https://osf.io/s8ec5/].	If the effect is not found, the effort moralization effect is **not replicated** 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.	Effort moralization theory's **generalizability** could be shown undetectable under the current conditions of the study.
	Individuals who invest **higher effort** in their **care work** are judged **higher in morality.**				If the effect is not found, it is potentially **not generalizable** to care work. Yet, the following analyses test the results in a more fine-grained manner.	Effort moralization is potentially not **generalizable** to the care context.
Aim 2: moral character as a function of gender, context, and effort						
Are there	Moral character	In study 1	Using **mixed-ANOVA** with 2	Based on the **smallest effect**	The ANOVA can	If effort doesn't

4

differences in effort moralization in the **work context** by gender and effort?	judgment differs by gender and effort.	(work context), we will sample *N* = **350** individuals (computed *N* = **324**)	(gender: female/male) x 2 (effort: high/low) design. **Gender** serves as a **between-subject** factor, and **effort** is a **within-subject** factor. **We further test the interaction of both terms.** The respective **Bayes Factor** is computed for each term.	**size of interest** (Lakens, 2022) approach, we aim to power for a small effect $\eta^2 = .01$ (Cohen, 1988) (*a* = .05, 1-β = .95, 2(gender)x2(effort)). This was computed, using G*Power 3.1.9.7 [see supplemental material, https://osf.io/s8ec5/].	**illustrate whether gender and/or effort** differentially influence **moral character** judgment in the **work context**.	affect moral character judgment, the effect is potentially not replicable in this context. If it shows differences by gender, the effect is potentially heterogeneous between genders (female/male).
Are there differences in effort moralization in the **care context** by gender and effort?	Moral character judgment differs by gender and effort.	In study 2 (care context), we will sample *N* = **350** individuals (computed *N* = **324**)	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. **We further test the interaction of both terms.** The respective **Bayes Factor** is computed for each term.	Based on the **smallest effect size of interest** (Lakens, 2022b) approach, we aim to power for a small effect $\eta^2 = .01$ (cite Cohen, 1988) (*a* = .05, 1-β = .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 **illustrate whether gender and/or effort** differentially influence **moral character** judgment in the **care context**.	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 heterogeneous between genders (female/male).
Aim 3: cooperation partner satisfaction as a function of gender and effort						
Do gender and effort influence	~~w~~**W**ork context: cooperation	In each study, we will	Using **mixed-ANOVA** with 2 (gender: female/male) x 2	Based on the **smallest effect size of interest** (Lakens,	We will be informed as to which degree **effort** is	Effort might not be a meaningful

Formatierte Tabelle

5

cooperation satisfaction?	satisfaction is predicted by gender and effort	sample sample $N = 350$ individuals (computed $N = 272$)	(effort: high/low) design. **Gender** serves as a **between-subject** factor, and **effort** is a **within-subject** factor. **We further test the interaction of both terms.** The respective **Bayes Factor** is computed for each term.	2022b) 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, https://osf.io/s8ec5/].	crucial for cooperation satisfaction **between women and men** in the **work context**.	predictor of cooperation satisfaction. Further, there might not be differences between females and males.
	eCare context: cooperation satisfaction is predicted by gender and effort		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. **We further test the interaction of both terms.** The respective **Bayes Factor** is computed for each term.		We will be informed as to which degree **effort** is crucial for cooperation satisfaction **between women and men** in the **care context**.	
Exploratory Analysis: are differences in effort moralization moderated by gender norm endorsement						
Are differences in effort moralization between genders moderated by gender norm	This will be tested in the **work** and **care** context	This exploratory analysis will be performed on the computed samples size	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	This exploratory analysis will be performed on the computed samples size of **Aim 2**	We will be informed whether gender norm endorsement moderates the influence of gender on effort moralization and whether the effect is generalizable on work	The effect is either generalizable on both contexts, context dependent or not observable with the present data.

Formatierte Tabelle

6

endorsement?		of **Aim 2**	commitment). **lmer-formula:** morality ~ gender*gender_norm + effort + (1|subject) We will further compare the **Bayes Factor** of the model against the model without gender norm endorsement.		and care contexts.	

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),~~, 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 (Brambilla & Leach, 2014; Goodwin et al., 2014; Wojciszke, 2005~~); Goodwin et al., 2014; Brambilla & Leach, 2014).~~. In this context, a particular bias has gained recent attention: the effort moralization effect (Amos et al., 2019; Bigman & Tamir, 2016; Celniker et al., 2023; 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 into 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)~~(Johnson & Park, 2021; Reed et al., 2007).

Interestingly, ~~the effort~~ the effort moralization effect persists~~.~~, even when ~~the added~~additional effort does not lead to increased performance ~~(e.g., better outcomes at work)~~ (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 (Mexiko: $d = .14$–$.28$, Germany: $d = .34$–$.37$, 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 career context.

Current gaps 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). Two-thirds 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 dollars 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~~found 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, cooperation partners are not always freely chosen but can be assigned as well (e.g., project assignments in the workplace). We, therefore, extend the literature by assessing individuals' satisfaction with assigned, instead of freely chosen, partners. This provides additional insights that reflect the cooperative dynamics frequently found in everyday life.

Gendered stereotyping in moral judgment and effort perceptions

As shown in prior research, social judgment is not immune to the influences of stereotyping, including gender biases. These extend to differing expectations of behavior and personality based on a person's gender. For instance, while men are often seen as more agentic, women are perceived as more communal ~~(e.g., caring or helpful) (Hentschel et al., 2019).~~ (e.g., caring or helpful; Hentschel et al., 2019). These expectations may inform differences in effort moralization~~, examplewise~~ for example, through backlashing~~,~~ and can differ between contexts.

This describes how expectations, for instance, those formed by gender, can lead to differing social judgments ~~(Rudman, 1998).~~ (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),~~ (Kennedy et al., 2016; Rudman, 1998), whereas men face greater criticism for non-agentic behavior in leadership contexts ~~(Moss-Racusin et al., 2010).~~ (Moss-Racusin et al., 2010).

See me, judge me, pay me

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.

The interplay between effort moralization and gender may expand to the caregiving contexts. Although gender roles have shifted, with more women entering the workforce (Toossi & Morisi, 2017)(Toossi & Morisi, 2017) and men contributing more to family labor (Sayer, 2016),(Sayer, 2016), women still do most of the care work (Charmes, 2019).(Charmes, 2019). These persistent gender role expectations shape how caregiving efforts are perceived. Research on double standards has shown that mothers face harsher criticism for low care efforts while fathers receive 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.

How can the present project inform psychological theorizing?

As previously outlined, multiple overarching theories may influence the moralization of effort across gender and contextual lines. The extant literature offers substantiation for several possible directions of effects, and our study may yield different outcome patterns. These include the following: (1) A gender-based discrepancy in the moral evaluation of effort is observed in the domains of work and care work, (2) This discrepancy is only evident in the work context, (3) This discrepancy is only evident in the care context, (4) No differences in the moral evaluation of effort are found based on the gender of the actors.

The first pattern (1) would indicate that the gender of the actors in question does play a role in the moralization of effort and that stereotypical gender role expectations have an influence on the moral judgments of others across contexts. Within this pattern, different

directions of effects are imaginable. For example, in the work context, gendered stereotypes could lead to a higher moral judgement of men, despite their level of effort, because their presence at work is stereotypically assumed and valued, while women are considered less capable to perform in the work context (Rudman, 1998; Sterling & Reichman, 2016). Conversely, the same stereotypes could potentially result in higher moral judgments for men in the high-effort condition but lower—compared to women—in the low-effort condition, as men are expected to demonstrate invested and agentic behavior at work. Consequently, demonstrating low effort could lead to lower moral judgments for men (Moss-Racusin et al., 2010).

Within the context of care, it is plausible that men will receive higher morality ratings in the high-effort condition compared to women, as the social expectations of women's pro-social behavior might render their high effort less exceptional. Consequently, even low-effort behavior exhibited by men might receive higher moral ratings than high-effort behavior displayed by women (Deutsch & Saxon, 1998). Conversely, while men might receive higher moral evaluation for high effort than women, low effort might be evaluated equally negatively across genders, as care work is expected and socially valued (Samtleben & Müller, 2022).

Naturally, it is also possible that the effects will be found in only one of the contexts eliciting the result patterns (2) or (3). This would suggest that the work or care context, respectively, is or has grown to be more resilient to stereotypical gender role expectations regarding effort moralization. For instance, differences may be found only in the care context, while gender stereotypes may not play a significant role for effort moralization in the work context (anymore). Alternatively, if no differences in effort moralization are identified in both contexts, as reflected by result pattern (4), the absence of effects could indicate neglectable influences of gender stereotypes on the moral judgment of effort.

12

Current study

The current project focused on three core aims-: I) Toto test the replicability of the effort moralization effect and explore its generalizability to the care work context-. II) Toto examine how the effort moralization effect interacts with gender across different contexts-. and III) Toto 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, 2022b), we powered both studies to detect a small effect (Cohen, 1988) in a 2X2 mixed ANOVA ($\eta^2 = .01$, $\alpha = .05$, $1-\beta = .95$). 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 (https://osf.io/s8ec5/)[1]. Participants were recruited via Prolific and consisted of US participantsindividuals based in the US.

Table 1

Descriptives by Study

	N	age: *M* (*SD*)	range	%female
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

---

[1] 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, while 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

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

See me, judge me, pay me

Table 2

Overview of Measures

Construct (*n* items)	Example item	Low anchor	High anchor
core goodness (6)[a]	honest	does not describe X well	describes X extremely well
value commitment (7)[a]	dedicated	does not describe X well	describes X extremely well
competence/warmth (2)[a]	competent	does not describe X well	describes X extremely well
effort (1)[b]	How much effort do you think X puts into his/her (care) work?	no effort at all	a lot of effort
quality (1)[c]	In your opinion, how well does X perform his/her (care) work?	very bad	very good
difficulty (1)[c]	Compared to other jobs/ care work, how difficult is X's (care) work.	not at all difficult	extremely difficult
work value (1)[c]	How valuable do you think X's (care) work is?	not valuable at all	extremely valuable
pay deservingness (work) (1)[a]	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	Imagine that X was paid for his/her	$0	$50

(care) (1)[a]	care work. How much should s/he be paid per hour?		
collaboration partner choice (1)[a]	[...] Please indicate how satisfied you would be to work with either X or Y?	extremely dissatisfied	extremely satisfied
gender role beliefs	Women with children should not work outside the home if they don't have to financially.	Strongly disagree	Strongly agree

Note. [a]These variables are the focal dependent measures, [b]This measure serves as manipulation check and exclusion criterion, [c]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[2]:

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

---

[2] 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).

See me, judge me, pay me

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. Subsequently, participants responded to the

items constituting the short version of the gender role belief scale (Brown & Gladstone,

2012). 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 all data ~~were~~was

collected via Prolific. Participants received compensation according to the platform's

standard rates [in Stage 2].

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 ($\frac{1}{7}$ x $\frac{1}{7}$ =) 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.

See me, judge me, pay me

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 low-effort conditions. We further computed the respective effect size (Cohen's d) and Bayes Factor (BF_{10}). 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) with an interaction term. For all terms, the respective Bayes Factor (BF_{10}) was computed to quantify evidence of absence and presence of effects.

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

19

See me, judge me, pay me

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 3

Comparison of Moral Character Judgment by Effort and Context

	Low effort: *M* (*SD*)	High effort: *M* (*SD*)	*d* [95% CI]	BF_{10}
		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

20

See me, judge me, pay me

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* ≥ .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), for which we further computed Bayes Factors.

Figure 1

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

See me, judge me, pay me
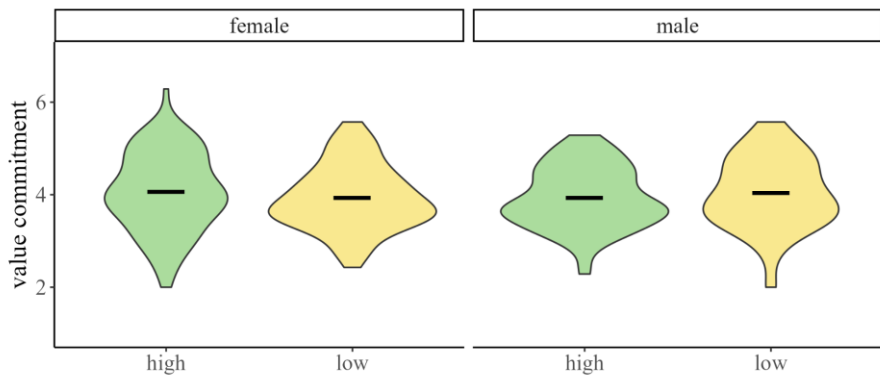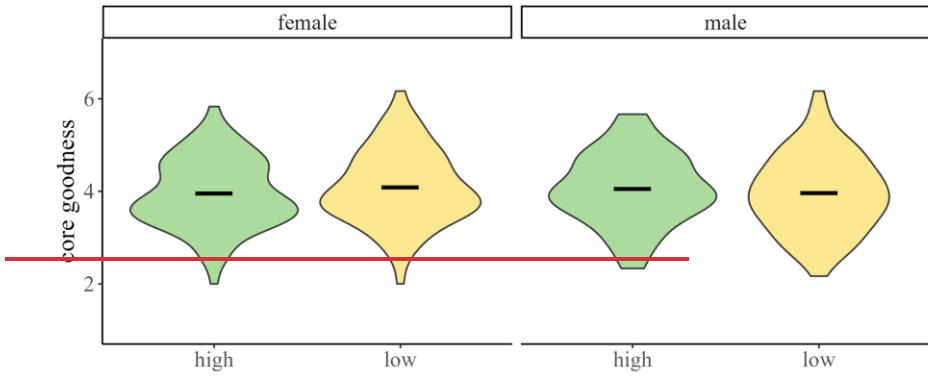
See me, judge me, pay me
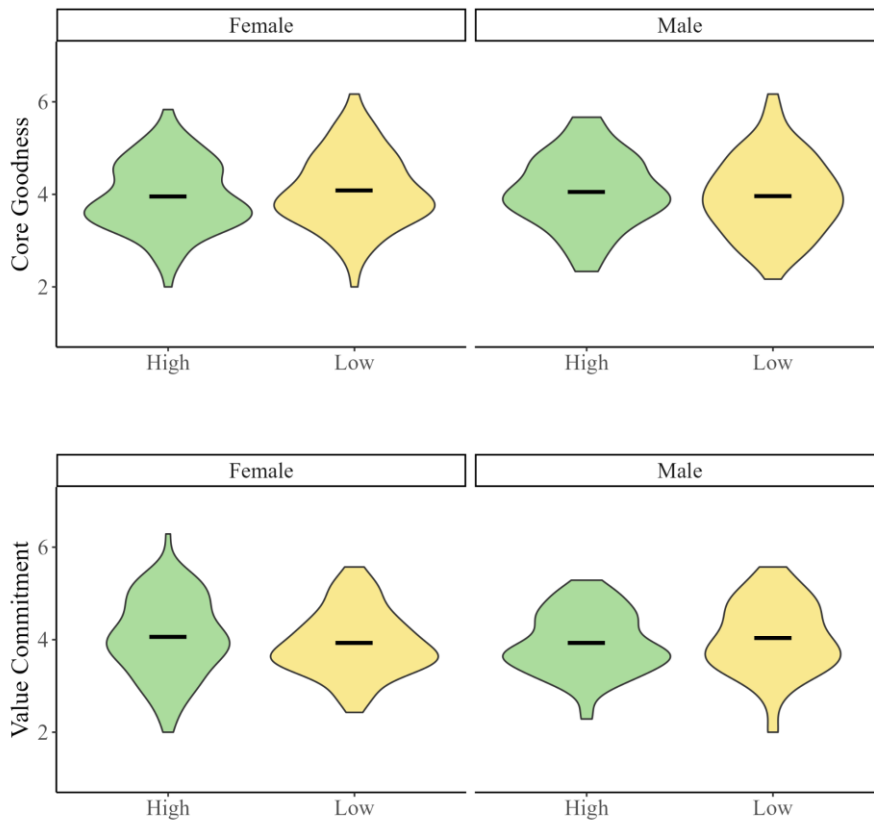
care context

      We fitted the 2x2 mixed ANOVA model (between: gender, within: effort), for which we further computed Bayes Factors.

Figure 2

Care Context: Mean Differences in Moral Judgment by Effort and Gender
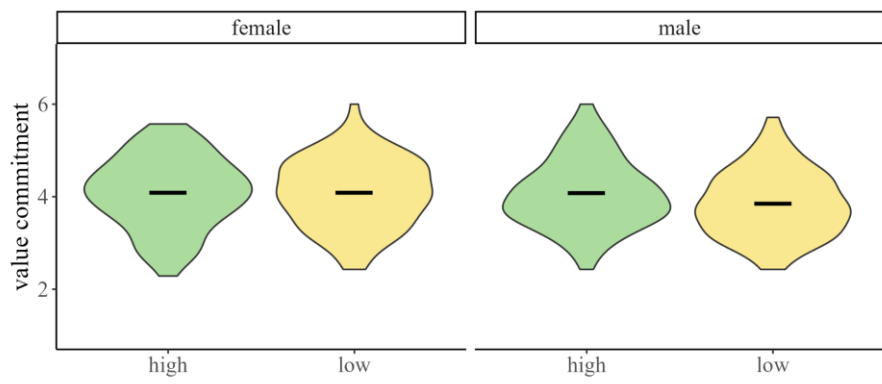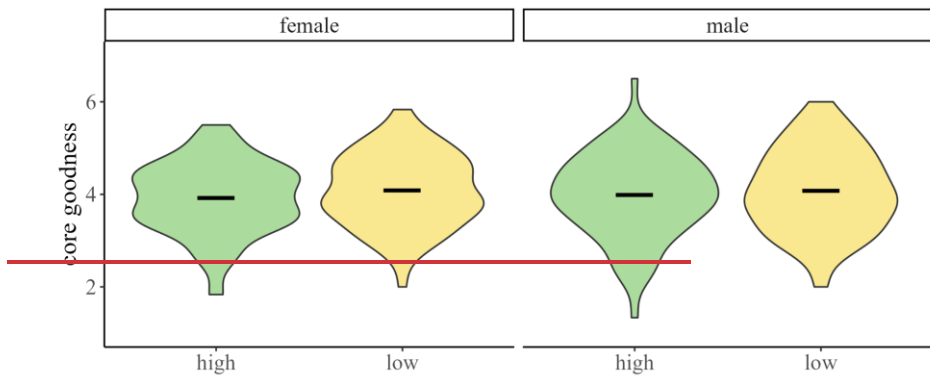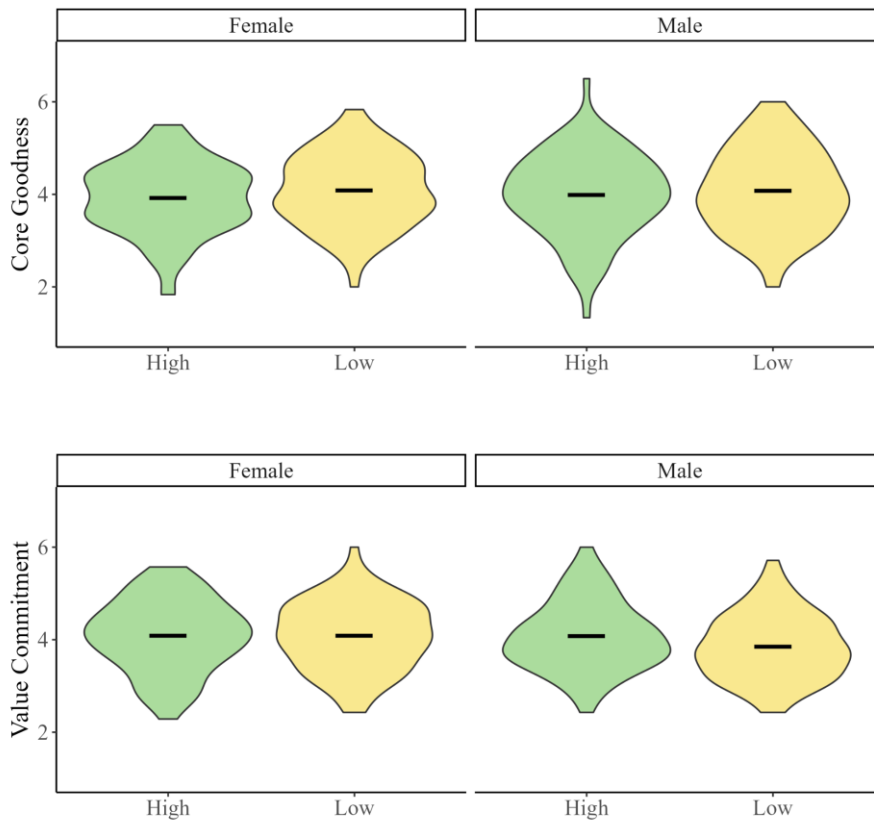
Further, we assessed suggested hourly salaries in both contexts to test for gender effects and whether effort differentially affected these between genders. [will be completed in Stage 2 manuscript].

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 a mixed-effect ANOVA to test the effect of effort (high effort/low effort) as a within-subjects factor and gender (female/male) as a between-subjects

See me, judge me, pay me

factor. The analysis was conducted by context (work/care). [will be completed in Stage 2

manuscript].

Work context

Figure 3

Work Context: Cooperation Partner Satisfaction
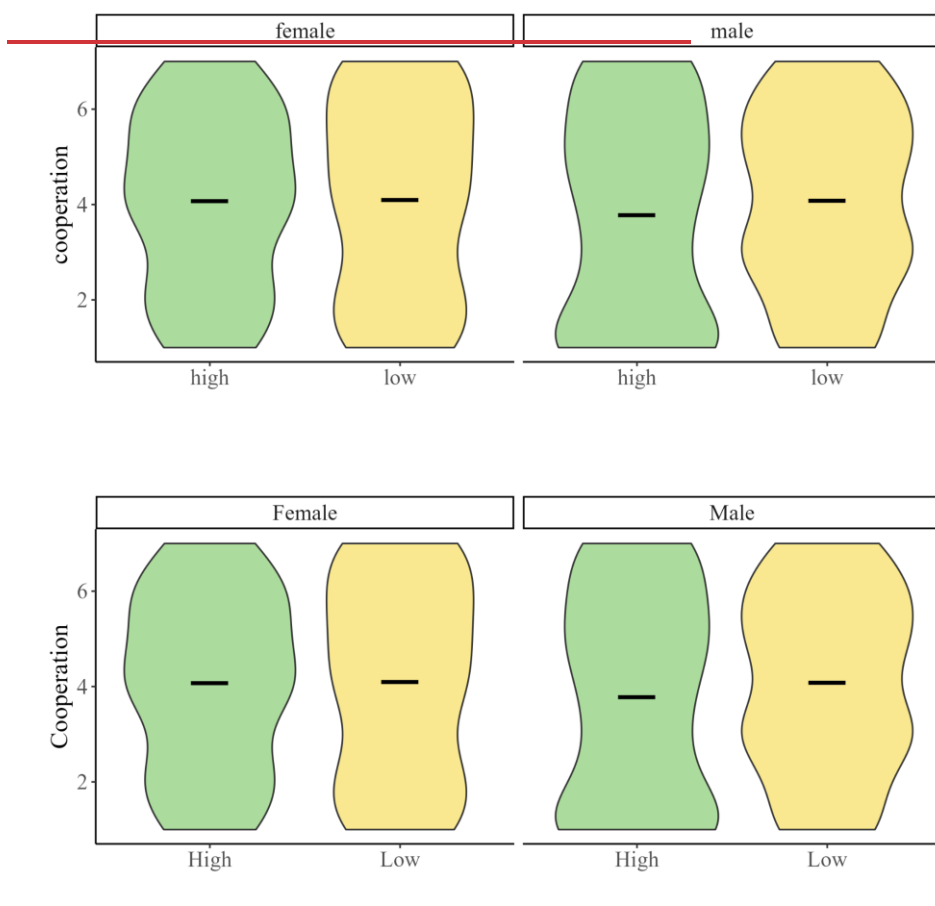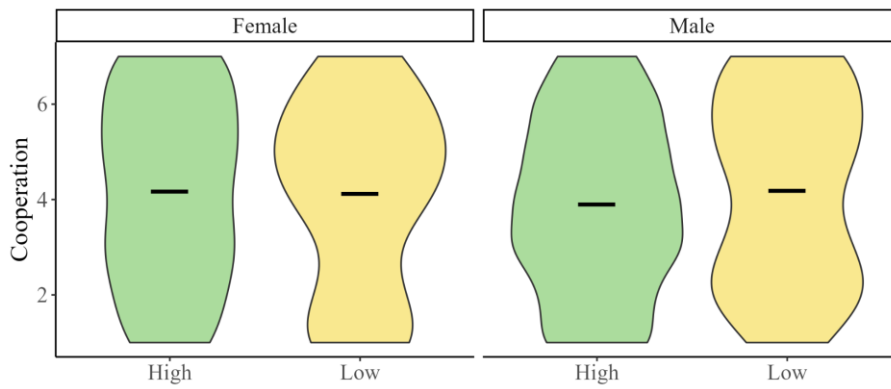
See me, judge me, pay me

Care context

Figure 4

Care Context: Cooperation Partner Satisfaction
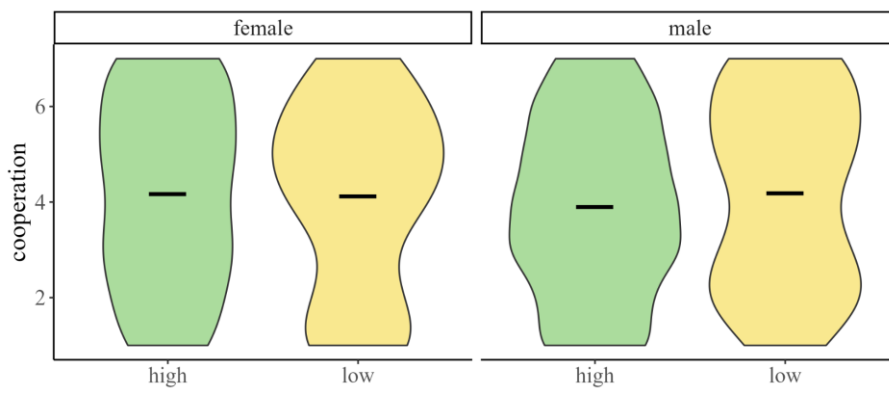
See me, judge me, pay me

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

Discussion

Summary

Limitations

Conclusion

Competing interests

The authors report no conflict of interest.

See me, judge me, pay me

Ethic approvement

The study was approved by the Departmental Review Board (DRB) of the Faculty of Psychology, Department of Occupational, Economic, and Social Psychology, University of Vienna (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. Amos, C., Zhang, L., & Read, D. (2019). Hardworking as a *European Journal of Social Psychology*h,euristic for *38*m(7), 1202–1217. https://doi.org/10.1002/ejsp.575oral

29

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.~~ 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.~~ unpaid care work-paid work connection. *SSRN Electronic 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. *Nature*, *456*(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.~~ 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.~~ short version of the gender role beliefs scale. *International Journal of Psychology and Behavioral Sciences*, *2*(5), 154–158.

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.~~ 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.~~ 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. 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. Hentschel, T., Heilman, M. E., & Peus, C. V. (2019). The *Psychology of Women Quarterly*m.ultiple *40*d(3), 353–363. https://doi.org/10.1177/0361684316634081imensions of

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. 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. gender raise the ethical bar? Exploring the punishment of ethical violations at work. *Academy of Management Proceedings*, *2016*(1), 11664. https://doi.org/10.5465/ambpp.2016.11664abstract

Lakens, D. (2022). Sample Size Justification. 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*Americans'. National Partnership for Women & Families. https://nationalpartnership.org/americans-unpaid-caregiving-worth-1-trillion-annually-women-two-thirds-work/*un*

Moss-Racusin, C., Phelan, J., & Rudman, L. (2010). When Men Break the Gender Rules: Status Incongruity and Backlash Against Modest Men. *paid 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-unpaid-caregiving-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~~New . ~~National Partnership for Women & Families. https://nationalpartnership.org/news_post/new-analysis-americans-unpaid-care-work-worth-more-than-1-trillion-each-year/~~*a*

~~Piazza, J., Goodwin, G. P., Rozin, P., & Royzman, E. B. (2014). When a Virtue is Not a Virtue: Conditional Virtues in Moral Evaluation.~~ *nalysis shows unpaid care work in the U.S. is worth more than $1 trillion each year*. National Partnership for Women & Families. https://nationalpartnership.org/news_post/new-analysis-americans-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.~~ 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.),~~ Samtleben, C., & Müller, K.-U. (2022). Care and careers: Gender (in)equality in unpaid care, housework and employment. *Research in Social Stratification and Mobility, 77*, 100659. https://doi.org/10.1016/j.rssm.2021.100659

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

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.~~ 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).~~ Sterling, J. S., & Reichman, N. (2016). Overlooked and ~~*U.S. female labor force participation rate 1990-2023*u.~~ ~~Statista. https://www.statista.com/statistics/191737/us-female-labor-force-participation-rate-since-1990/~~ndervalued: Women in

~~Tissot, T. T., & Roth, L. H. O. (2024).~~ p~~*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*~~rivate ~~. https://osf.io/5tkfr~~l

~~Toossi, M., & Morisi, T. L. (2017).~~ aw ~~*BLS Spotlight on Statistics: Women in the Workforce Before, During, and after the Great Recession*~~p. ~~https://hdl.handle.net/1813/78334~~ractice.

~~Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.~~ *Annual Review of Law and Social Science*, *12*(Volume 12, 2016), 373–393. https://doi.org/10.1146/annurev-lawsocsci-120814-121705

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 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