Received: 26 February 2020 | Revised: 10 August 2020 | Accepted: 10 August 2020

DOI: 10.1002/bdm.2208

RESEARCH ARTICLE

WILEY

Does she compensate the victim while he punishes the perpetrator? No gender differences in anonymous economic games across 11 nations

Angela R. Dorrough[1] | Maria I.T. Olsson[2] | Laura Froehlich[3] | Andreas Glöckner[1,4] | Sarah E. Martiny[2]

[1]Social Cognition Center Cologne, Department of Psychology, University of Cologne, Cologne, Germany

[2]Research Group Social Psychology, Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway

[3]Department of Social Psychology, Faculty of Psychology, FernUniversität in Hagen, Hagen, Germany

[4]Max Planck Institute for Research on Collective Goods, Bonn, Germany

Correspondence
Angela R. Dorrough, Social Cognition Center Cologne, Department of Psychology, University of Cologne, Richard-Strauss-Str. 2, Cologne 50931, Germany.
Email: angela.dorrough@uni-koeln.de

Funding information
University of Hagen

Abstract

Social role theory posits that occupational gender roles give rise to gender differences in behavior, such that men and women engage in qualitatively different prosocial behaviors. Therefore, we expected that women who observed an unfair situation (involving a victim and a perpetrator) would respond by demonstrating communal prosocial behavior (by compensating the victim), whereas men would respond with agentic prosocial behavior (by punishing the perpetrator). Furthermore, on the basis of social role theory, we expected that gender differences would be more pronounced in countries with a more unequal distribution of men and women in communal and agentic occupational roles. The current research tested the predictions using an economic game. Two studies consisting of samples from 10 countries (Study 1, N = 1,791) and a student sample from Germany (Study 2, N = 193) showed no support for the predicted gender differences in prosocial behavior and no systematic relationship between prosocial behavior and gender roles across countries.

KEYWORDS

agentic, communal, gender differences, helping, prosocial behavior, social role theory

1 | INTRODUCTION

Prosocial behavior is associated with helping, sharing, comforting, rescuing, and defending another individual, but it can also involve supporting the collective, group, or nation (Eagly, 2009). Previous research shows that men and women differ in the degree to which they engage in prosocial behavior. However, the direction of this gender difference is less clear. For instance, some research demonstrates that females are more prosocial than males (e.g., Carlo, Roesch, Knight, & Koller, 2001; Russell, Hart, Robinson, & Olsen, 2003; see also the meta-analysis by Fabes & Eisenberg, 1998), whereas other studies have shown that males are more prosocial than females (e.g., Eagly & Crowley, 1986). In the present research, we investigate whether women and men exhibit *different kinds* of prosocial behavior.

1.1 | Gender differences in prosocial behavior

We argue that there is no simple answer to the question of which gender is more prosocial. Instead, on the basis of social role theory, we argue that men and women engage in different kinds of prosocial behavior (Eagly, 2009). According to the theory, behavioral gender differences are driven by an unequal distribution of men and women in different social roles (e.g., Eagly, Wood, & Diekman, 2000), which shape expectations regarding gender-typed behavior. Across the world, women more frequently occupy communal, caring-oriented social roles, whereas men more frequently occupy agentic, achievement-oriented social roles (Kan, Sullivan, & Gershuny, 2011). When observing men and women in these roles, people infer that they have specific traits (i.e., communal traits on the part of women vs. agentic traits on the part of men) that qualify them for these roles. As men and women are internally motivated to meet these expectations (Greenwald et al., 2002) and avoid backlash for engaging in gender stereotype-incongruent behavior, men are more likely to engage in agentic (prosocial) behaviors (e.g., behavior that can increase the status of the helper), whereas women are more likely to engage in communal (prosocial) behaviors (e.g., empathy-related behavior; Eagly, 2009).

1.1.1 | Gender differences using social psychological paradigms

To investigate this assumption, Eagly (2009) reviewed research on gender differences in prosocial behavior in a range of social situations that involved interacting with strangers, romantic partners, family members, and work colleagues and concluded that men and women are not *more or less* prosocial. Instead, Eagly noted that women seem more likely to engage in communal (i.e., relational, friendly, compassionate, and comforting) prosocial behaviors, whereas men seem more likely to engage in agentic (i.e., dominant, assertive, aggressive, and risky) prosocial behaviors. For example, women show more prosocial behavior in dyadic relationships: they are more likely to empathize and engage in self-disclosure than men (e.g., Rose & Rudolph, 2006) and are more likely to provide emotional support in marriages (Neff & Karney, 2005). Men, on the other hand, are more likely to offer help to strangers (Johnson et al., 1989) and to help in the presence of others (e.g., De Caroli & Sagone, 2013).

1.1.2 | Gender differences using economic paradigms

Reviewing the economic game literature, which uses money transfers as a measure of prosocial behavior, results in the same conclusion; that is, a more differentiated perspective on gender differences is necessary. Balliet, Li, Macfarlan, and Van Vugt (2011) conducted a meta-analysis of 272 effect sizes for gender differences in the amount men and women transferred to an interaction partner. They found no overall gender differences but concluded that gender differences in prosocial (transfer) behavior were game and context dependent. Indeed, a closer look into the literature reveals that gender differences in economic game behavior seem to hinge on a specific game feature: Women usually act more prosocially in games requiring an altruistic motive for transfer behavior, whereas men transfer more money to their interaction partners when the game involves an element of risk. For example, in dictator games, in which the player assigned to be a "dictator" can decide whether they wish to selflessly share their experimental endowment with a passive player, women often transfer more money to the interaction partner than men do (e.g., Eckel & Grossman, 1998; Engel, 2011). However, in economic games (e.g., trust games, Romano, Balliet, Yamagishi, & Liu, 2017; prisoner's dilemma games, Dorrough & Glöckner, 2019) that involve the risk of exploitation (where the player transfers money and the interaction partner does not), men typically transfer more money to interaction partners than women do (for an overview regarding the different motives in economic games, see Thielmann, Böhm, & Hilbig, 2015).

In sum, previous research using social psychological as well as economic paradigms suggests that men and women are not more or less prosocial but rather prosocial in different ways. The assertion that men and women engage in different kinds of prosocial behavior (in line with gender role expectations), however, has only been inferred from comparing results between studies using different research designs (Eagly, 2009). A direct experimental test of this assertion within one study is still missing. The current project aims to fill this gap to get a better understanding of gender differences in prosocial behavior and to bring findings from previous research together. Taken together, based on the theoretical and empirical considerations above, it can be predicted that men are more likely to engage in agentic forms of prosocial behavior, whereas women are more likely to engage in communal forms of prosocial behavior.

1.1.3 | Gender differences in prosocial behavior across countries

In addition to investigating whether men and women engage in different kinds of prosocial behavior, our research also tests whether gender differences in prosocial behavior vary between countries. According to social role theory, observing men and women in different (occupational) roles shapes gender stereotypes, which, when internalized, gives rise to gender differences in behavior in the direction of the stereotype (e.g., Eagly & Wood, 2012). As countries differ in occupational gender segregation, we should observe larger gender differences in behavior in countries with more occupational gender segregation (Eagly & Wood, 1999). Previous cross-national research on mating preferences and gender equality supports this assumption: Gender differences in preferences for stereotype-congruent partners were less pronounced in countries that ranked high (rather than low) on a gender equality index (Eagly & Wood, 1999). In line with social role theory and previous findings, it can be predicted that gender

differences in prosocial behavior increase as gender segregation in occupational roles increases. However, it has to be noted that results from some previous multinational studies are contrary to this assumption. For example, gender differences in preferences (e.g., social and risk preferences; Falk & Hermle, 2018) and basic values (e.g., power, hedonism; Schwartz & Rubel-Lifschitz, 2009) were more pronounced in countries that ranked high (rather than low) on gender equality. Moreover, research has shown variation in the size of gender differences in prosocial behavior (in a prisoner's dilemma game) across different countries, without finding a systematic relationship between gender differences and gender equality (as measured by the Gender Inequality Index [GII]; http://hdr.undp.org/en/content/gender-inequality-index-gii; Dorrough & Glöckner, 2019). Thus, it seems especially important to shed light on the question of whether or not gender segregation in the occupational domain influences gender differences in prosocial behavior.

1.2 | The present research

The present research composes of two studies. In both Studies 1 and 2, we use an economic game to experimentally investigate whether men and women engage in different kinds of prosocial behavior. Furthermore, in Study 1, we test whether gender differences in prosociality are influenced by a country's level of gender equality.

1.2.1 | Compensation/punishment game

One way to measure prosocial behavior in a standardized and controlled way is through economic games. Economic games are widely used by economists and psychologists in mononational as well as multinational studies (e.g., Dorrough & Glöckner, 2016; Romano et al., 2017). One advantage of using economic games over self-report is that economic games capture actual incentivized behaviors rather than mere intentions that might be biased by social desirability (see Camerer, 2003; Thielmann, Heck, & Hilbig, 2016). Economic games model complex social interactions by concentrating on a few central dependent variables and actual behavior. Games are conducted anonymously and provide a consistent study environment for all participants. Thus, in contrast to field studies or other types of experiments, economic games facilitate control over endogenous and exogenous factors (Murnighan & Wang, 2016), ensuring internal validity. Economic games are particularly suitable for multinational research as they provide only minimal context information, reducing the threat of poor cross-cultural validity. Furthermore, the anonymity of the game interactions rules out potential confounds, such as the extent to which participants perceive interaction partners to be attractive or similar to themselves. In economic games, participants receive an initial endowment and are required to decide how much of that endowment they would like to transfer to an interaction partner. The present research employs a third-party compensation/punishment game (e.g., Leliveld, Van Dijk, & Van Beest, 2012; Weng, Fox, Hessenthaler, Stodola, & Davidson, 2015), which allows for two potential prosocial behaviors. Specifically, participants observe that a person receives an unfair monetary transfer from a third player. They can then use their own monetary resources to *compensate* the victim (i.e., the player who has received an unfair transfer) in the form of a money transfer or to *punish* the perpetrator (i.e., the player who has made the unfair transfer) by reducing their payoff. As the participant has to sacrifice some of their own endowment without financial benefit to compensate and punish, both behaviors are forms of prosocial behavior (i.e., altruistic punishment and altruistic compensation; Fehr & Fischbacher, 2003).

Communal behaviors involve empathy, caring for others, sociability, and affection (e.g., Abele, Uchronski, & Suitner, 2008). We consider compensation transfers to be a display of communal prosocial behavior as these transfers are relational and comfort the victim. In support of this claim, Weng et al. (2015) found in a compensation/punishment game that empathic concerns promoted compensation of a victim but not punishment of a perpetrator.

Agency is commonly associated with aggression, protectiveness, and striving for power and status (e.g., Abele et al., 2008). Research suggests that punishment transfers are a display of agentic prosocial behavior. First, punishment is driven by spontaneous anger reactions (Mischkowski, Glöckner, & Lewisch, 2018) and is often perceived as an act of aggression by the punished person, leading to acts of revenge (e.g., counter-punishment; Nikiforakis, 2008). Second, punishment may result in behavioral change on the part of the perpetrator (e.g., Glöckner, Kube, & Nicklisch, 2018), which prevents the perpetrator from continuing to victimize others. As such, punishment may be an act of protection because it communicates to perpetrators that their behavior violates a norm, is not acceptable, and should be changed. Third, evidence indicates that punishment is related to power and dominance: Participants who see themselves as more powerful engage in more punishing transfers in different economic games (Chierchia, Parianen Lesemann, Snower, Vogel, & Singer, 2017). Further work demonstrates that participants make more punishment transfers in economic games with unstable power hierarchies (Dorrough, Glöckner, & Lee, 2017). These results show that the relevant dimensions of agency are activated not only in direct interpersonal settings with long-term interactions but also in anonymous economic games with short-term interactions. This might partially be due to individuals overgeneralizing from social contexts to such games (Rand et al., 2014).

In summary, in the present research, we consider compensation a communal type of prosocial behavior and punishment an agentic type of prosocial behavior. Our paradigm minimized additional motives for third-party transfers such as demand characteristics and reputation by ensuring that interactions were anonymous and that participants interacted with their interaction partners only once and were not directly affected by the unfair transfer. To the best of our knowledge, our studies are the first to investigate gender differences in prosociality by providing male and female participants with a communal and an agentic behavioral option. Applying the above-mentioned reasoning to game behavior, we hypothesize that

men will make more punishment transfers (H1a), whereas women will make more compensation transfers (H1b). We additionally hypothesize that gender differences in compensation and punishment transfers will increase as gender segregation in occupational roles increases (H2).

2 | OVERVIEW OF STUDIES

The present research consists of two studies. In Study 1, participants from 10 countries took part in the compensation/punishment game. During the game, participants (who had been allocated to the role of "Player E") indicated decisions for transfers of Talers, which was the experimental currency, to other players. Participants made a decision after being informed about a fair transfer (where a player had transferred 50 out of 100 Talers1 to another player) and an unfair transfer (where a player had transferred 0 out of 100 Talers to another player). For each transfer, participants indicated how they would like to invest their own endowment (50 Talers). Although participants were asked to indicate decisions for transfers in response to fair and unfair behavior, only the behavioral response for the unfair treatment is relevant to our research questions, as we were only interested in prosocial behavior (helping a person in need) elicited by the unfair condition. Thus, only data from this scenario will be analyzed.2 Participants could choose to spend their endowment by (a) keeping all of their Talers to themselves, (b) transferring some or all of their Talers in the form of punishment to the perpetrator (i.e., the player who behaved unfairly), and/or (c) transferring some or all of their Talers in the form of compensation to the victim (i.e., the player who was treated unfairly). Participants were informed that Talers transferred as punishment would reduce the perpetrator's experimental payoff, whereas Talers transferred as compensation would increase the victim's experimental payoff. To test our assumption that gender differences would increase with (perceived) occupational gender segregation, participants indicated to what extent a range of agentic and communal occupations were segregated by gender (see below).

To ensure the generalizability of our findings, Study 1 was conducted not only in Western, educated, industrialized, rich, and democratic (WEIRD; Henrich, Heine, & Norenzayan, 2010) but also in non-WEIRD countries such as Chile and Indonesia. Study 2 serves as a conceptual replication of Study 1 but includes a more convenient student sample from a single country (Germany). Our hypotheses were preregistered on the Open Science Framework3 (Study 1, https://osf.io/q5a4v/ Study 2, https://osf.io/p5vyk/). All data, analysis scripts, and material files (English version) can be accessed at https://osf.io/re7n3/. All translations of the scales can be accessed4 at https://osf.io/7ybns/.

Study 1 was approved by the ethics committee of the University of Göttingen and was conducted in accordance with the ethical guidelines of the German Psychological Association (DGP) and the American Psychological Association (APA). Study 2 was a conceptual replication of Study 1 with only minor changes to the measures and procedure.

3 | STUDY 1: SAMPLES FROM 10 COUNTRIES

3.1 | Method

3.1.1 | Participants and design

Data were collected in the summer of 2018. We employed a 10 (country) × 2 (participant gender: male vs. female) × 2 (victim gender: male vs. female) mixed design with victim gender as a within-participant factor. Participants were recruited via an online panel provider (Toluna: https://de.toluna.com/). Our second hypothesis required sufficient variance in gender segregation in occupational roles, so we selected countries that varied in their rankings on the GII (http://hdr.undp.org/en/content/gender-inequality-index-gii), which is partly determined by the proportion of women in the labor force. We used data from 2017, which were the most recent data at the time of study planning. For pragmatic reasons, we did not consider countries with small Toluna panels (<50,000) or more than one official language (e.g., Paraguay and Belgium). This procedure resulted in the selection of the following countries (sorted by increasing gender equality): Indonesia (GII rank 104 of 160), Colombia (rank 87), Mexico (rank 76), Chile (rank 72), Russia (rank 53), the United States (rank 41), China (rank 36), Japan (rank 22), Spain (rank 15), and Sweden (rank 3). In each country, the sample was recruited to be representative of the general population in terms of age and gender.

Study 1 is part of a larger project that collected data over two time points (the data analyzed here were measured at the second time point/Part 2).5 The a priori calculated sample size (n = 208 per country) was based on the design for the superordinate project using an effect size estimate of d = 0.14 for gender differences in game behavior (see Dorrough & Glöckner, 2019). To be able to detect small effects, the required sample size was estimated at 200 participants (for Players E, i.e., those who indicated decisions for compensation and punishment transfers) with a desired power of 80%. This power analysis was conducted with the use of G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). For the analysis, we assumed a repeated measures mixed analysis of variance (ANOVA) as the closest pragmatic approximation for the cluster-corrected regression analysis we intended to run, as no convenient method for power estimation of the latter analysis was available. We aimed to achieve this sample size in each country. However, because of participant dropout, some countries have lower sample sizes. Thus, results for individual countries must be interpreted with caution.

For the analyses, and as registered a priori on the Open Science Framework, we included only participants who (a) completed both Parts 1 and 2, (b) entered a valid participant code, and (c) indicated a country of origin that corresponded to their country selection when registering with the panel provider. From the final data set, we excluded participants who indicated different countries of origin in Parts 1 and 2 or who indicated a gender other than male or female (n = 3). In addition, we excluded participants (n = 112) who were (randomly) assigned to the role of "perpetrator" (Player C) or "victim"

(Player D) rather than the role of interest (i.e., "observer"; Player E). The final Player E sample is listed in Table 1. It has to be noted that because of exclusion based on the above-mentioned criteria and because of dropouts between Parts 1 and 2, some national subsamples (for Players E) are smaller than originally planned. Sensitivity analyses assuming repeated measures mixed ANOVAs (see above) for the different countries show that the effects that can be detected with the available national samples (with a power of .80; alpha = .05) range from d = 0.14 (Russia; n = 216) to d = 0.21 (the United States; n = 114).

It took participants approximately 45 min to complete Parts 1 and 2. In addition to receiving a basic payment, participants received a bonus payment that ranged from the equivalent of 0–4 US dollars and that was based on one randomly selected incentivized task from Part 1 or 2. For example, the bonus payment could be based on one decision in the compensation/punishment game by the participants themselves (if they had been allocated to the role of Player E or C) and/or their interaction partner(s). Participants' payment was first calculated in "Talers," which was converted to the respective national currency for payout. The basic payment was credited to participants' Toluna accounts directly after they completed the study. The bonus payment, which depended on the decision of the participant and/or their interaction partner(s), was calculated after study completion in that we matched participants in groups of three according to their player roles (one Player C, one Player D, and one Player E). The bonus payment for these three players was determined by Player C's transfer and the compensation and punishment decision of Player E for this transfer. The bonus payment was credited to the participants' Toluna accounts by the Toluna project manager.

The study was translated from English into the language of each subsample by a professional translation agency (https://www.e-kern.com/). Thereafter, researchers in psychology (who were fluent in the language of each translated survey) checked the translations against the English version. If necessary, the surveys were sent back to the translation company that considered the language edits made by the researchers. The final versions were subsequently checked by a native speaker; again, if necessary, minor changes were made following their feedback.

TABLE 1 Subsamples in Study 1

Country	*N*	Age range in years	% female
Chile	142	18–82	46
China	163	20–75	45
Colombia	184	18–71	52
Indonesia	174	18–69	56
Japan	204	20–81	50
Mexico	194	18–75	51
Russia	216	19–77	58
Spain	198	18–78	49
Sweden	202	18–86	53
United States	114	19–86	54

Note: This table shows the subsamples that entered our analyses. That is, only Role E players are depicted in this table and considered in our power considerations and sensitivity analyses.

3.1.2 | Materials and procedure

Participants first read information about data protection and details of the study (i.e., duration and payment) and then provided informed consent. At the beginning of the study (both in Parts 1 and 2), participants were informed about the calculation of their bonus payment. Participants learned that they would receive a bonus payment based on their answers and/or the answers of other participants in a randomly selected task.

Gender segregation in occupational roles

Participants reported the proportion of men across a range of communal and agentic occupations in their country: "In your country, what proportion (%) of individuals who work as [occupation] are male?" On the basis of the previous research (Koenig & Eagly, 2014), we assessed the perceived proportion of men across the following occupations: nursery teachers, geriatric aides, nurses, secretaries, and therapists (communal occupations) as well as police officers, attorneys, CEOs, rescue service workers, soldiers, politicians, and judges (agentic occupations).

Compensation/punishment game

Second, participants took part in the compensation/punishment game. Participants were randomly assigned to one of three different player roles—C, D, or E. Only the decisions of Player E were relevant to the hypothesis testing and analyzed. Even though data from Players C and D were not required to test our hypotheses, a minority of participants were assigned to those roles (n = 112). Doing this allowed us to calculate participants' bonus payments based on participants' *actual* behaviors and thereby avoid deception in terms of how the bonus payment was calculated.6 Player C had to decide how much of their endowment of 100 Talers they wanted to transfer to the recipient (Player D). The amount not transferred was kept by Player C. Player C was reminded that the amount they kept might form the basis of their bonus payment. Player C could transfer either 50 or 0 Talers to the recipient (Player D). Player D had a passive role and could not make a decision in the compensation/punishment game. Player E was informed that they could transfer Talers as compensation to Person D (which would increase Player D's payoff) and as punishment to Person C (which would reduce Player C's payoff). Player E was asked to make transfer decisions for two different scenarios (using a variant of the strategy method): In the first scenario, Player E was asked how to invest their Talers if Player C makes a fair transfer and sends 50 Talers to Player D ("How much of your Talers do you want to transfer to Person C and Person D if Person C keeps 50 Talers for herself [himself] and transfers 50 Talers to Person D?"). In the second scenario, Player E was asked how to invest their Talers if Player C makes an unfair transfer and sends 0 Talers to Player D ("How much of your Talers do you want to transfer to Person C and Person D if Person C

keeps 100 Talers for herself [himself] and transfers 0 Talers to Person D?"). Only the latter (unfair) transfer was examined in our analyses. For each of the two decisions, Player E received an endowment of 50 Talers and was asked to decide how to invest them. Player E could divide their Talers by (a) keeping all of their Talers, (b) transferring Talers in the form of punishment to Player C, and/or (c) transferring Talers in the form of compensation to Person D. To make prosocial behavior more profitable, Player E was informed that the Talers they transferred as either punishment or compensation would be tripled by the experimenter (for a similar approach, see Leliveld et al., 2012).

To avoid confounds because of cultural differences in how much men and women interact with members of the opposite gender in their daily lives, participants were informed that they were interacting with players of the same gender. For exploratory analyses, we added an additional round of the game in which participants were informed that they were interacting with a victim (Player D) of the opposite gender.

Following the procedure of previous multinational research (Romano et al., 2017), we limited the amount of written information and used pictorial instructions to illustrate the rules of the game (see Figure 1 for an example). Participants were able to start the game once they had correctly answered three comprehension questions (e.g., "Person C transfers 0 Talers to Person D. How many Talers do you have to transfer to Player D for them to receive 30 Talers as compensation?"). At the end of the study, participants were provided with contact information that they could use to request additional information about the study.

3.2 | Results

As preregistered, for our main analyses, we only included data from same-gender interactions (in which the gender of Player D matched the gender of Player E; see reasoning for that decision above). Overall, the descriptive results did not indicate that men and women engage in different kinds of prosocial behavior (Figure 2). In fact, in all countries, men and women used punishment and compensation in a similar manner (Figure S1) in that both men and women invested more Talers in compensation (men: M = 14.15, SD = 12.33; women: M = 13.22, SD = 11.35) than in punishment (men: M = 9.05, SD = 10.34; women: M = 8.61, SD = 9.48; see Table S1 for country level results). These findings speak against our assumption of gender-typed prosociality.

In order to determine the statistical significance of gender differences (or lack thereof) across different prosocial behaviors, we ran an ordinary least squares (OLS) regression analysis with cluster-corrected standard errors at the participant level to account for dependencies in error terms. Specifically, we predicted transfer behavior by prosocial dimension (punishment vs. compensation) and participant gender controlling for country differences by including country dummies. This analysis did not reveal the hypothesized interaction between participant gender and prosocial dimension (Table 2, Model 1).7 Predicting transfer behavior for the different dimensions separately (punishment in Model 2; compensation in Model 3) to test H1a and H1b more directly did not reveal significant effects of participant gender.8 When including data from interaction partners of the opposite gender and controlling for whether Player D had the same gender as Player E (yes vs. no) in an exploratory analysis, the interaction between participant gender and prosocial dimension remained nonsignificant (b = −0.49, t(1790) = −0.80, p = .423). The three-way interaction was also nonsignificant (b = 0.08, t(1790) = 0.11, p = .914). Finally, analyses per country showed no significant interaction effect between participant gender and prosocial dimension in any of the 10 countries (all p > .232). All these analyses speak for the robustness of the null findings.

We predicted that gender differences in communal and agentic prosocial behavior would increase as people perceived more gender segregation in communal and agentic occupations (H2). We calculated an index of gender segregation of labor by subtracting the perceived proportion of men in communal occupations from the perceived proportion of men in agentic occupations. The index could vary from 100 (100% men in agentic professions and 0% men in communal professions) to −100 (0% men in agentic professions and 100% men in communal professions). In line with Koenig and Eagly (2014), the index of gender segregation of labor was positive (M = 24.73, SD = 23.37), suggesting that participants perceived men as more likely to hold agentic occupations than communal occupations. On average, participants indicated that 68% (43%) of people working in agentic (communal) professions were men. However, as expected, we found

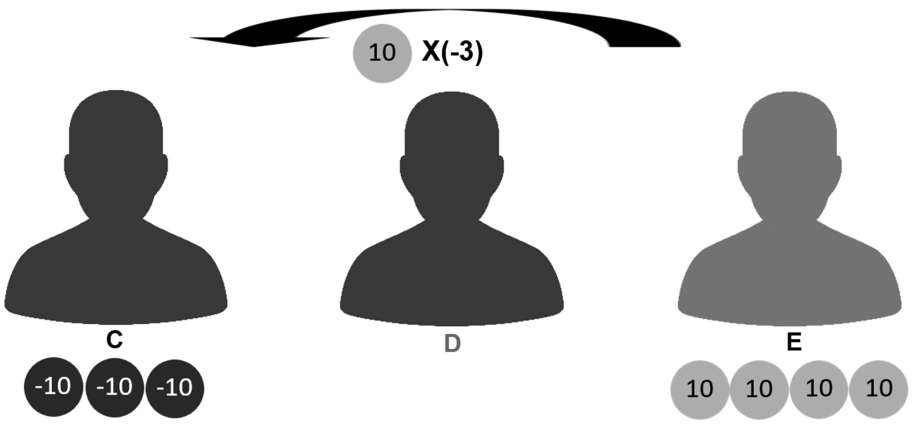

FIGURE 1 This figure presents an example illustration describing the punishment rule of the compensation/punishment game (i.e., if Player E transfers 10 Talers in the form of punishment to Player C, 30 Talers are deducted from Player C's account)

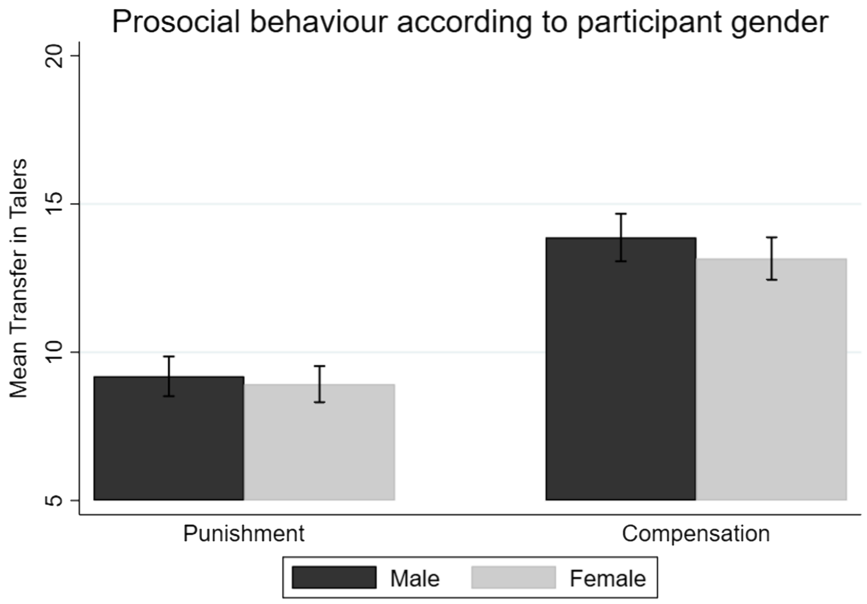

FIGURE 2 Prosocial behavior measured as amount transferred by dimension (punishment vs. compensation) and participant gender in Study 1. Spikes represent 95% confidence intervals based on cluster-corrected standard errors. The *p* value refers to the interaction between participant gender and dimension. A graphical illustration for each country can be found in Figure S1) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Transfer behavior in Study 1

	(1)		(2)		(3)	
	Transfer	95% CI	Transfer	95% CI	Transfer	95% CI
Participant gender (female = 0; male = 1)	0.650 (1.18)	[−0.43, 1.73]	0.194 (0.42)	[−0.71, 1.10]	0.662 (1.20)	[−0.42, 1.75]
Dimension (compensation = 0; punishment = 1)	−4.236*** (−9.12)	[−5.15, −3.32]				
Participant gender * dimension	−0.445 (−0.62)	[−1.84, 0.95]				
Constant	13.19*** (35.87)	[12.47, 13.91]	8.975*** (28.31)	[8.35, 9.60]	13.17*** (35.10)	[12.43, 13.90]
Observations	3,582		1,791		1,791	
Participants/clusters	1,791		1,791		1,791	
Adjusted R^2	0.042		0.005		0.000	

Note: Transfer behavior is predicted by participant gender, dimension, and their interaction (Model 1) as well as by gender for both dimensions separately (punishment in Model 2 and compensation in Model 3). Country dummies are not reported. The table contains unstandardized coefficients. *t* statistics are in parentheses.
***$p < .001$.

that the perceived segregation of labor varied across countries, from $M = 18.40$ in Japan to $M = 31.43$ in Russia, which suggests that occupations had a more equal distribution of women and men in some countries than in others (for more details, see Froehlich, Olsson, Dorrough, & Martiny, 2020). To test H2, we aggregated the division of labor index at the country level. We then ran a multilevel regression analysis predicting prosocial behavior by dimension (L1), participant gender (L1), the interaction of dimension and participant gender, as well as the cross-level interaction of the division of labor index (L2) with the other predictors (group-mean centering used for Level 1 predictors, grand-mean centering used for the Level 2 predictor; see Enders & Tofighi, 2007). The degree of between-participant and between-country variability in the mean level of prosocial behavior is captured by the Level 1 (*SD* intercept = 1.32, 95% CI [0.28, 6.28]) and Level 2 (*SD* intercept = 0.28, 95% CI [0.03, 3.01]) random parameter estimates of the intercept, respectively. The Level 2 random parameter estimates of the division of labor index denote whether there is between-country variability in gender differences in compensation and punishment transfers (*SD* slope = 0.00, 95% CI [0.00, 2.83]). As with the results reported above, we did not observe a simple effect for participant gender, $b = 0.43$, $z = 1.18$, $p = .238$, but we did find a significant effect of prosocial dimension on transfer, $b = -4.45$, $z = -12.57$, $p < .001$. Contrary to H2, we did not observe the hypothesized cross-level interaction, $b = -0.03$, $z = -0.15$, $p = .882$. We ran an exploratory analysis where we replaced the gender segregation of labor index with other indicators of gender equality. However, the

conclusions regarding the interaction term remain unchanged when using the 2018 GII (b = −4.66, z = −0.90, p = .368) and with women's labor force participation as part of the GII (b = 0.05, z = 0.39, p = .700).

3.3 | Discussion

Study 1 investigated whether women and men engage in different kinds of prosocial behavior as a function of gender roles in the occupational domain. In sum, data from Study 1 supported neither the prediction of gender-differentiated prosociality (H1a and H1b) nor the prediction of larger gender differences in countries that have more (perceived) gender segregation in communal and agentic occupations (H2). However, we found a substantial simple effect of dimension (see Figure 2 and Table 2) in that participants, irrespective of their gender, preferred to compensate rather than punish.

Exploratory analyses revealed that the null findings replicated when the victim was the opposite gender from the participant. One potential limitation of Study 1 is that participants might have perceived the punishment option as less attractive than the compensation option. We observed an unexpected significant effect of dimension, in that participants of both genders used an average of 4 Talers less (of an endowment of 50 Talers; β = −0.19, p < .001) to punish (i.e., engage in agentic prosocial behavior) than to compensate (i.e., engage in communal prosocial behavior). Thus, we cannot exclude the possibility that a potential gender difference was masked by the fact that the punishment option was less attractive, and therefore, both women and men made little use of it. To rule out this possibility, we adapted our paradigm in two ways to make the punishment option more attractive in Study 2.

4 | STUDY 2: STUDENT SAMPLE REPLICATION

To test the robustness of the null effect concerning gender differences in communal and agentic prosocial behavior, we conducted a conceptual replication of Study 1 within a single cultural context (Germany). Study 2 included some alterations to the compensation/punishment game designed to increase the likelihood of punishment being used. To reduce comparisons between the compensation and the punishment transfer (which could lead to a preference for the compensation transfer as a more socially acceptable kind of prosocial behavior), participants in Study 2 were asked to make decisions about punishment and compensation transfers in separate rounds. We expected that this procedure would lead to a higher use of the punishment option (for a similar procedure, see Raihani & Bshary, 2015). In addition, the effectiveness of punishment (i.e., the ratio with which punishment points reduced other players' payoffs) was increased in one condition to make this behavioral option more attractive.

4.1 | Method

4.1.1 | Participants and design

Data of N = 193 participants (18–64 years of age, 51% female)9 were collected in the on-site laboratory of the University of Cologne in February 2019 using a 2 (participant gender: male vs. female) × 2 (punishment effectiveness: medium vs. high) × 3 (behavioral option: compensation vs. punishment vs. both) × 2 (victim gender: male vs. female) between-within design (the last two factors were varied within participants). The sample size was based on the same power calculations that informed the sample size for Study 1. Participants received a basic payment of 3 Euros and a chocolate bar in exchange for taking part in the study. Participants' bonus payment for the compensation/punishment game was calculated in the same way as in Study 1. Participants could pick up their bonus payment on two predetermined dates at different university locations. The study took approximately 15 min to complete. After completing the study, the Taler income was converted into Euros using a conversion rate of 100 Talers = 6 Euros. On average, participants earned 8.70 Euros (approx. 9.80 US dollars) for participating in the study.

4.1.2 | Materials and procedure

Study 2 followed the same procedure as Study 1. The compensation/punishment game in Study 2 was identical to the one in Study 1 except for the changes described below. These changes were implemented to allow for a strong test of gender-differentiated prosocial behavior by increasing the overall use of punishment behavior (in order to reduce potential bias toward compensation behavior). The first change we implemented was separating the punishment and compensation rounds (see Raihani & Bshary, 2015 for a similar procedure) to prevent participants from comparing the relative value of punishment and compensation (and therefore being more likely to opt for the more socially desirable option, which would be compensation). The second change we implemented was to add a second condition with a higher efficiency of punishment to make this behavioral option more attractive. In the *medium punishment efficiency condition* (Condition 1), participants first played separate compensation and punishment rounds (presented in randomized order). In the compensation round, participants could only keep their Talers and/or use them to compensate the victim. In the punishment round, participants could only keep their Talers and/or use them to punish the perpetrator. Afterwards, participants played a round in which they could use their Talers to keep, punish, and/or compensate (replicating the design of Study 1). Punishment efficiency in all three rounds was the same as in Study 1 (1:3). The second, *high punishment efficiency condition* (Condition 2), which was implemented to make the punishment transfer more attractive, followed the same procedure with the exception that punishment efficiency was increased to 1:5 (i.e., each punishment Taler reduced the outcome of the other player by a factor of 5).

4.1.3 | Results

As in Study 1, only Player E's transfers in the unfair scenario were analyzed.10 To achieve maximum power, choices from all three rounds were included, that is, one punishment decision and one compensation decision from the first two rounds as well as one punishment and one compensation decision from the last round. The results pointed in the expected direction, in that men were more inclined to punish, whereas women were more inclined to compensate (see Figure 3). Averaged across all rounds and conditions, men used slightly fewer Talers for compensation (men: M = 12.37 [SD = 10.50]; women: M = 12.97 [SD = 8.65]) and slightly more Talers for punishment (men: M = 10.64 [SD = 10.53]; women: M = 9.07 [SD = 10.48]) compared with women. Descriptive statistics for all rounds and conditions are provided in Table S3.

In accordance with the preregistered analysis plan, we ran an OLS regression with cluster-corrected standard errors at the participant level predicting transfer by participant gender (female = 0; male = 1), dimension (compensation = 0; punishment = 1), and their interaction. We controlled for round (i.e., sequence), punishment efficiency, and gender of Player D (different gender = 0; same gender = 1). We did not find the hypothesized interaction between participant gender and prosocial dimension (see Table 3, Model 4). This was also true when we controlled for the two-way interactions between punishment efficiency and gender, punishment efficiency and dimension, and the three-way interaction (t = 1.52, p = .130). When predicting transfer for the two dimensions separately (punishment in Model 5, compensation in Model 6) to directly test H1a and H1b, we did not observe significant effects of participant gender on transfer.

When analyzing the medium punishment efficiency condition (Table S2, Model 2) and the high punishment efficiency condition (Table S2, Model 3) separately, no significant interaction effects were observed. The replication condition (Table S2, Model 4) with medium efficiency, in which participants could choose to use their Talers to keep, compensate, and/or punish, also did not reveal an interaction effect (see Table S3). When analyzing the punishment-only and compensation-only rounds separately in this condition, women and men did not differ with regard to their transfer behavior for punishment (b = 3.59, p = .072) or for compensation; (b = 0.20, p = .906). Although the simple effect of dimension was somewhat smaller than in Study 1, we found that participants, irrespective of their gender, preferred the communal prosocial behavior to the agentic prosocial behavior option (see Table 3).

For a maximally powerful test of our main hypotheses, we conducted an overall analysis combining both studies (total N = 1,984) predicting transfer by gender, dimension, and their interaction while controlling for whether Player D (i.e., the victim) had the same gender as the participant. Again, we did not find a significant interaction, b = −0.01, t(1983) = −0.01, p = .990. A sensitivity analysis showed that with this sample size, small effects (f = .10) could be detected with a statistical power of .95.

5 | GENERAL DISCUSSION

This research tested the hypothesis that men and women engage in qualitatively different prosocial behaviors. Eagly (2009) hypothesized that because of gender roles in the labor market, women are inclined to engage in communal (e.g., empathic, caring, sociable, and affectionate) forms of prosocial behavior, whereas men are inclined to engage in agentic (e.g., aggressive, protective, and striving for power and status) forms of prosocial behavior. On the basis of social role theory, we also expected that gender differences would be particularly pronounced in countries where men and women are perceived to be unevenly distributed across communal and agentic occupations (e.g., Eagly & Wood, 1999). The results of our studies did not support either of our hypotheses. We did not find evidence for the hypothesis that men and women engage in different kinds of prosocial behavior.

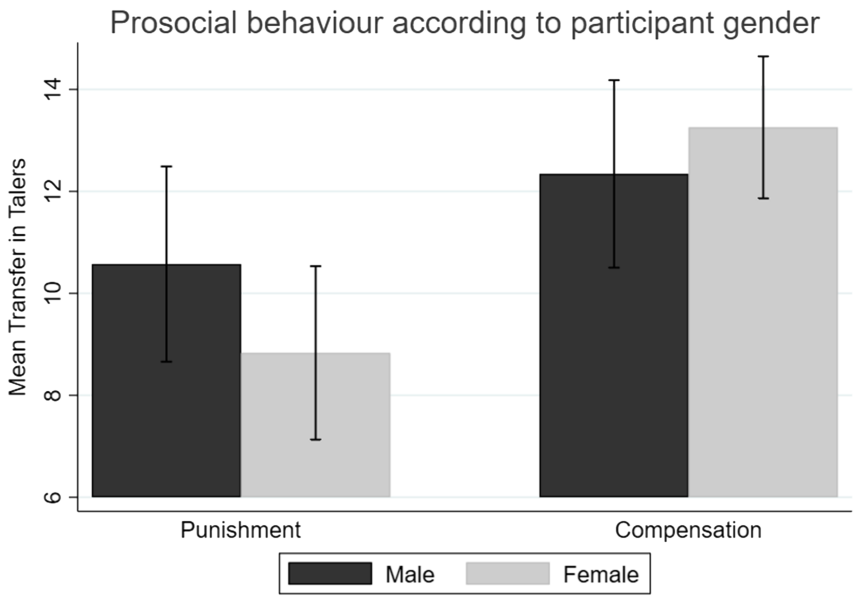

FIGURE 3 Gender differences in amounts transferred by scenario (punishment vs. compensation) in Study 2. Spikes represent 95% confidence intervals based on cluster-corrected standard errors [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Transfer behavior in Study 2

	(4)		(5)		(6)	
	Transfer	95% CI	Transfer	95% CI	Transfer	95% CI
Participant gender (female = 0; male = 1)	−0.614 (−0.54)	[−2.87, 1.64]	1.552 (1.21)	[−0.98, 4.09]	−0.610 (−0.53)	[−2.86, 1.64]
Dimension (compensation = 0; punishment = 1)	−3.042** (−2.88)	[−5.12, −0.96]				
Participant gender * dimension	2.169 (1.33)	[−1.04, 5.38]				
Round 1	2.614*** (4.08)	[1.35, 3.878]	2.614*** (4.08)	[1.35, 3.878]		
Round 2	4.326*** (7.32)	[3.16, 5.49]			4.326*** (7.32)	[3.16, 5.49]
Punishment efficiency (medium efficiency = 0; high efficiency = 1)	−1.152 (−1.27)	[−2.93, 0.63]	−1.381 (−1.07)	[−3.91, 1.15]	−0.923 (−0.81)	[−3.17, 1.32]
Constant	11.41*** (14.03)	[9.80, 13.01]	8.626*** (7.91)	[6.478, 10.778]	11.15*** (13.09)	[9.47, 12.83]
Observations	1,544		772		772	
Participants/clusters	193		193		193	
Adjusted R^2	0.053		0.020		0.049	

Note: Transfer behavior in Study 2 predicted by participant gender, dimension, and their interaction (Model 4) as well as by gender for both dimensions separately (punishment in Model 5 and compensation in Model 6). Round dummies were included with Round 3 as the reference category. Effects of Player D gender are omitted. The table contains unstandardized coefficients. *t* statistics are in parentheses.
**$p < .01$.
***$p < .001$.

Contrary to our first hypothesis, in most countries, men do not seem to be more likely than women to choose the agentic behavioral option (see Table S1). Only in the US and the Chilean subsamples did we observe a substantial difference between men and women in the expected direction. Likewise, in none of the participating countries did women seem to be more likely than men to choose the communal behavioral option. In fact, our results suggest that both men and women show a preference for communal over agentic prosocial behavior. This seemed to be the case despite our efforts to maximize the utility of punishment transfers in Study 2. One reason could be that punishment, in contrast to compensation, results in an overall collective loss, as the money used for punishment does not benefit anyone in monetary terms. Contrary to our second hypothesis, we did not find a relationship between gender differences in prosocial behavior and gender roles in the labor market, as measured by the perceived distribution of men and women in different communal and agentic occupations. Moreover, gender differences in prosocial behavior did not substantially vary across countries that ranked differently on gender equality. Hence, our data speak against a robust effect of gender-typed prosocial behavior that is contrary to social role theorizing and previous research showing gender differences in economic preferences (e.g., Croson & Gneezy, 2009; Falk & Hermle, 2018) and in prosocial transfer behavior (e.g., Eckel & Grossman, 1998; Engel, 2011). Furthermore, contrary to previous research, our data did not indicate a positive (Eagly & Wood, 1999) or negative (Falk & Hermle, 2018) relationship between gender equality and gender differences. Given these null findings, one should be cautious making strong general claims concerning gender differences in prosocial behavior. More differentiated perspectives are required.

5.1 | Potential reasons for null findings and future directions

The overall null findings observed in our research could reflect the fact that men and women do not differ in the kind of prosocial behaviors they engage in. This finding is in line with the gender similarity hypothesis (Hyde, 2005) that men and women are psychologically more similar than different and differ only in very few aspects (e.g., motor behavior and physical aggression). It is also possible that gender differences in prosocial behavior exist but that gender differences are very small and not very robust. To explore this, future research could use multiple measures (instead of only one behavioral measure) for the communal and the agentic prosocial domain. Multiple measures would allow for aggregating behaviors in the respective domains and may provide a more accurate estimate of gender differences by reducing noise in the data. To increase test power to detect cross-level effects, future studies could test our predictions on a larger sample (preferably 30–50 nations; Maas & Hox, 2005).

Another potential reason for the observed null findings is that gender differences in prosocial behavior are very context sensitive (see also Balliet et al., 2011 for a similar conclusion). One advantage of using economic games is that only very little contextual information is provided, which excludes potential confounding factors. However, this can also be a disadvantage as we may inadvertently remove the very factors that trigger gender differences in real life. Although in Study 1, we did not find gender differences in the type of prosocial behavior men and women engage in, there was a (non-significant) trend in the predicted direction in Study 2. Specifically, women tended to compensate the victim more than men, whereas men

tended to make more use of punishment than women did. Thus, it is possible that gender differences in behavior are more pronounced in lab studies (e.g., because of more direct interaction with the interaction partners of both genders being in the same room), highly educated samples, or specific countries (e.g., more egalitarian countries; Falk & Hermle, 2018). Furthermore, the compensation/punishment game provides two behavioral options that differ with regard to only one aspect (i.e., providing comfort vs. behaving in a more dominant and aggressive manner). Future studies could test whether our findings replicate when participants are provided with more contextual information (e.g., more information about their interaction partners, more information about the situation where help is required, such as whether there are bystanders or not) and more behavioral options that do not (only) concern monetary help as is the case in economic games. For example, future research could assess gender differences in prosocial behavior in the presence versus absence of a bystander and toward a group versus an individual (Eagly, 2009). It is possible that these factors may ignite gender differences in prosocial behavior (Eagly, 2009). Furthermore, future research could investigate gender differences across different settings, for example, in the leisure/recreational domain as well as in the work domain. Identifying moderating factors could generate knowledge about under which conditions gender differences in behavior are likely to occur and thus provide boundary conditions for the predictions of social role theory.

Finally, the null findings could be explained by the processes expected to underlie the relationship between gender roles and gender-differentiated behaviors. According to social role theory, the observation of men and women in different social and occupational roles leads to gender stereotypes, which in turn result in gender differences in behavior, cognition, and affect. The relationship between gender stereotypes and gender-differentiated outcomes is (among other factors) mediated by social regulation (i.e., adhering to the expectations other people have of men and women; Eagly & Wood, 2012). This means that even if people observe gender segregation, this may not necessarily affect their *own* behavior if they are not motivated to adhere to these expectations. Future research could include measures of participants' expectations regarding men's and women's prosociality and the degree to which they are motivated to act in accordance with these expectations. This could also add to knowledge about the generalizability of social role theory.

5.2 | Strengths and limitations

In the current research, we addressed the hypothesis that men and women engage in qualitatively different prosocial behavior using a highly controlled economic game paradigm. Economic games offer several advantages (e.g., *actual* behavior is measured; potentially confounding factors, such as reputation concerns, are precluded), especially in a cross-national context (because of high cross-cultural validity). We included both WEIRD and non-WEIRD subsamples in Study 1 and a convenient student sample in Study 2 to ensure the generalizability of our findings. To the best of our knowledge, this research is the first that has experimentally investigated gender differences in prosociality by providing male and female participants with a communal and an agentic behavioral option in response to observing unfair treatment (i.e., a situation where prosocial behavior is warranted). Furthermore, by sampling from countries in Study 1 that varied in inequality according to the GII (http://hdr.undp.org/en/content/gender-inequality-index-gii), we provided a comprehensive comparison of gender differences in prosocial behavior across countries that vary in gender segregation in occupational roles. Despite some strengths, the present research has several limitations that we outline below for future researchers to consider. One potentially suboptimal design feature in the present research is that participant gender might have become salient prior to/when interacting with other players in the compensation/punishment game. In Study 1, participants were asked to estimate the proportion of men in different occupations before taking part in the economic game. Moreover, in both studies, the gender of the victim was manipulated within participants such that participants always interacted with a same-gender victim before interacting with an opposite-gender victim. These design features may have made the purpose of the study fairly obvious, which in turn may have resulted in demand effects (Nichols & Maner, 2008). However, for various reasons, we think that it is unlikely that our null findings are due to demand effects. First, the study involved real incentives, and behavior in line with demands would have been costly and less likely than in merely hypothetical studies. Second, at the end of Study 1 (Study 2), only 1.7% (4.1%) of participants made a gender-related comment in their feedback on the study, which suggests that participants were not aware they were participating in a study on gender differences. It should be noted, however, that we did not explicitly ask participants to speculate about the purpose of the study. Anyhow, previous research shows that when group identities are made salient, participants are more likely to show attitudes (e.g., Steele & Ambady, 2006) and behavior (e.g., Spencer, Steele, & Quinn, 1999) consistent with stereotypes for this group. Thus, it can be argued that making gender salient should have increased rather than suppressed gender differences. The fact that we did not find gender differences despite potentially making gender salient speaks in favor of the null hypothesis. Nevertheless, future studies may opt to pose gender-related questions after participants have completed the economic game or at a separate time point. Additionally, future research could assess transfer behavior toward other players that have not had their gender specified.

Another potential limitation of the present research is that it did not include simultaneous, direct interactions between players. Instead, participants indicated transfer behavior for various scenarios (i.e., the perpetrator makes a fair vs. unfair transfer) using a variant of the strategy method (i.e., participants indicate decisions/their strategy for different roles or scenarios), which is widely used in economic and psychological research. It was only after study completion that participants were matched with other participants in the study and their bonus payment was calculated based on their behavior and/or the behavior of their interaction partner(s). Brandts and Charness (2011) compared the use of the direct response (i.e., participants interact

with each other in real time) with the strategy method (i.e., participants indicate decisions for different scenarios) in economic games. They concluded that whereas the behavior of the person that makes the first decision (e.g., the trustor in the trust game) is very similar across both methods, the behavior of the person who acts in reaction to the behavior of other people (e.g., the punisher or trustee in a trust game) can differ. The authors argue that the difference stems from emotional responses that should be stronger in "hot" (i.e., direct) as compared to "cold" (i.e., strategy method) decisions. This could explain why the employment of the strategy method (vs. direct response) does not seem to influence third party punishers who make punishment decisions in reaction to observed behavior (Jordan, McAuliffe, & Rand, 2016), a finding especially relevant to the present research. However, whether this is also the case for third-party compensation and whether these differences are moderated by participant gender are open questions that should be addressed in future research. Using a direct response method in future studies has the additional advantage that participants might be less skeptical as to whether they are actually interacting with other participants, which may have influenced their transfer behavior. Although we could not find any indication of such skepticism in the feedback participants provided at the end of the survey, we cannot completely rule out that participants treated the compensation/punishment game as a hypothetical interaction making our monetary incentivization ineffective.

5.3 | Conclusion

On the basis of the findings from Studies 1 and 2, we conclude that gender differences in communal and agentic prosocial behavior—if they exist at all—are small and not robust, in line with the similarity hypothesis that men and women are more similar than different (Hyde, 2005). According to our a priori power analysis, our sample sizes were large enough to detect small to medium effects. Future studies could increase statistical power in order to detect *very* small gender differences in behavior. This is especially relevant given that the majority of gender differences in the cognitive and social domain are in the close-to-zero or small range (Hyde, 2005). Furthermore, future studies could enrich the context and add additional measures of agentic and communal prosocial behavior. On the basis of our results, one should be careful not to make strong claims concerning gender differences in the usage of different kinds of prosocial behavior. Such claims might cause harm in reinforcing perceptions that persons of a specific gender are unfit for specific occupations (lack of fit model; e.g., Heilman, 1983; Heilman & Caleo, 2018). Specifically, women might be given fewer opportunities to enter high-status positions (Heilman & Caleo, 2018) where an agentic form of prosociality is expected. On the other hand, men might be disadvantaged in care-giving jobs (e.g., Cejka & Eagly, 1999) where a communal form of prosocial behavior is expected. Furthermore, the expectation of large gender differences can induce self-esteem problems and influence couple communication and conflict (for a discussion, see Hyde, 2005). In sum, on the basis of our research presented in this article, we conclude that men and women do not substantially differ concerning the type of prosocial behavior they show in economic games. Further research is needed to assess gender-typed prosociality in other domains and contexts.

DATA AVAILABILITY STATEMENT

All data, analysis scripts, and material files (English version) can be accessed at https://osf.io/re7n3/.

ACKNOWLEDGEMENTS

This research was funded by the internal gender research funding program of the University of Hagen.

ORCID

Angela R. Dorrough https://orcid.org/0000-0002-5645-949X

ENDNOTES

[1] 100 Talers = 100 US cents.

[2] Recording participants' responses following the fair transfer was required for the calculation of their bonus payment.

[3] For Study 1, additional hypotheses for behavioral intentions were specified. These hypotheses are discussed in a manuscript available from the authors upon request.

[4] The translation of other measures belonging to different projects are also provided here.

[5] In Part 1, we recorded participants' social and risk preferences. Part 1 also included some items of the World Value Survey. In Part 2, we recorded participants' transfer behavior in a prisoner's dilemma game and behavioral intentions in the work context. Because participants did not receive feedback during the study, we did not expect carry-over effects from Study 1 to the compensation/punishment game.

[6] The number of participants in a given country that received a bonus payment dependent on behavior in the compensation/punishment game was determined by the player role with the smallest number of participants. For example, if we had 10, 15, and 163 participants allocated to Roles C, D, and E, respectively, 10 participants in each player role received a bonus payment dependent on their and/or their interaction partners' decision(s) in the compensation/punishment game. All other participants from this country received a bonus payment dependent on another randomly determined incentivized task from Part 1 or 2. The participants whose payment was determined by the compensation/punishment game were matched (each Player E was matched with one Player C and one Player D). Depending on Player C's decision, the bonus payment was calculated. For example, if Player C made a fair transfer, Player E's decision for a fair transfer was used. To give a further example, if player C chose an unfair transfer of 0 and Player E decided to transfer 0 Talers in the form of punishment to Player C and 10 Talers in the form of compensation to Person D, Player C received a bonus payment of 100 Talers, Player D received a bonus payment of 3 * 10 Talers (the compensation), and Player E received a bonus payment of 40 Talers (50–10 Talers).

[7] A multilevel linear random effects regression with two levels (participant and country level) and random intercepts predicting transfer by participant gender, prosocial dimension, and their interaction (both variables group-mean centered) led to the same conclusion ($b = -0.47$, $z = -0.65$, $p = .513$, for the interaction).

[8] In the preregistration of Study 1, we specified a slightly different approach. Rather than absolute transfers, we planned to use transfers relative to the overall amount that people transferred. When using this index instead of the absolute values, conclusions remain unchanged.

[9] One participant indicated a gender different from male or female. Three participants stated an age <18. These participants were excluded from the analyses.

[10] In contrast to Study 1 and to achieve more observations per participant, we preregistered that we would include decisions for interaction partners of the opposite gender and control for this factor in the analyses. Results were unchanged when we restricted the models reported in Table 2 to interactions with the same gender.

REFERENCES

Abele, A. E., Uchronski, M., & Suitner, C. (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

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, *137*(6), 881–909. https://doi.org/10.1037/a0025354

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*, 375–398. https://doi.org/10.1007/s10683-011-9272-x

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Russell Sage.

Carlo, G., Roesch, S. C., Knight, G. P., & Koller, S. H. (2001). Between- or within-culture variation? Culture groups as a moderator of the relations between individual differences and resource allocation preferences. *Applied Developmental Psychology*, *22*, 559–579. https://doi.org/10.1016/S0193-3973(01)00094-6

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, *25*(4), 413–423. https://doi.org/10.1177/0146167299025004002

Chierchia, G., Parianen Lesemann, F. H., Snower, D., Vogel, M., & Singer, T. (2017). Caring cooperators and powerful punishers: Differential effects of induced care and power motivation on different types of economic decision making. *Scientific Reports*, *7*, 11068. https://doi.org/10.1038/s41598-017-11580-8

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474. https://doi.org/10.1257/jel.47.2.448

De Caroli, M. E., & Sagone, E. (2013). Self-efficacy and prosocial tendencies in Italian adolescents. *Procedia-Social and Behavioral Sciences*, *92*, 239–245. https://doi.org/10.1016/j.sbspro.2013.08.666

Dorrough, A. R., & Glöckner, A. (2016). Multinational investigation of cross-societal cooperation. Proceedings of the National Academy of Sciences of the United States of America, *113*(39), 10836–10841. https://doi.org/10.1073/pnas.1601294113

Dorrough, A., Glöckner, A., & Lee, B. (2017). Race for power in public good games with unequal, unstable punishment power. *Journal of Behavioral Decision Making*, *30*(2), 582–609. https://doi.org/10.1002/bdm.1976

Dorrough, A. R., & Glöckner, A. (2019). A cross-national analysis of sex differences in prisoner's dilemma games. *British Journal of Social Psychology*, *58*, 225–240. https://doi.org/10.1111/bjso.12287

Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist*, *64*, 644–658. https://doi.org/10.1037/0003-066X.64.8.644

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*(3), 283–308. https://doi.org/10.1037/0033-2909.100.3.283

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, *54*(6), 408–423. https://doi.org/10.1037/0003-066X.54.6.408

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories in social psychology*, *2*. Sage Publications. https://doi.org/10.4135/9781446249222.n49

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). Taylor & Francis Group

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men: Evidence from dictator experiments. *The Economic Journal*, *108*(448), 726–735. https://doi.org/10.1111/1468-0297.00311

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Fabes, R. A., & Eisenberg, N. (1998). *Meta-analysis of age and sex differences in children's and adolescents' prosocial behavior*. Unpublished manuscript. Department of Family Resources & Human Development, Arizona State University.

Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, *362*(6412), 1–6. https://doi.org/10.1126/science.aas9899

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–791. https://doi.org/10.1038/nature02043

Froehlich, L., Olsson, M. I. T., Dorrough, A. R., & Martiny, S. E. (2020). Gender at work across nations: Men and women working in male-dominated and female-dominated occupations are differentially associated with agency and communion. *Journal of Social Issues*. Advance Online Publication. https://doi.org/10.1111/josi.12390

Glöckner, A., Kube, S., & Nicklisch, A. (2018). The joint benefits of observed and unobserved social sanctions. *Journal of Behavioral and Experimental Economics*, *75*, 105–116. https://doi.org/10.1016/j.socec.2018.05.007

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25. https://doi.org/10.1037//0033295X.109.1.3

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. *Research in Organizational Behavior*, *5*, 269–298.

Heilman, M. E., & Caleo, S. (2018). Combatting gender discrimination: A lack of fit framework. *Group Processes & Intergroup Relations*, *21*(5), 725–744. https://doi.org/10.1177/1368430218761587

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Johnson, R. C., Danko, G. P., Darvill, T. J., Bochner, S., Bowers, J. K., Huang, Y.-H., … Pennington, D. (1989). Cross-cultural assessment of altruism and its correlates. *Personality and Individual Differences*, *10*(8), 855–868. https://doi.org/10.1016/0191-8869(89)90021-4

Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, *19*, 741–763. https://doi.org/10.1007/s10683-015-9466-8

Kan, M. Y., Sullivan, O., & Gershuny, J. (2011). Gender convergence in domestic work: Discerning the effects of interactional and institutional barriers from large-scale data. *Sociology*, *45*, 234–251. https://doi.org/10.1177/0038038510394014

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371–392. https://doi.org/10.1037/a0037215

Leliveld, M. C., Van Dijk, E., & Van Beest, I. (2012). Punishing and compensating others at your own expense: The role of empathic concern on reactions to distributive injustice. *European Journal of Social Psychology*, *42*, 135–140. https://doi.org/10.1002/ejsp.872

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Mischkowski, D., Glöckner, A., & Lewisch, P. (2018). From spontaneous cooperation to spontaneous punishment—Distinguishing the underlying motives driving spontaneous behavior in first and second order public good games. *Organizational Behavior and Human Decision Processes*, *149*, 59–72. https://doi.org/10.1016/j.obhdp.2018.07.001

Murnighan, J. K., & Wang, L. (2016). The social world as an experimental game. *Organizational Behavior and Human Decision Processes*, *136*, 80–94. https://doi.org/10.1016/j.obhdp.2016.02.003

Neff, L. A., & Karney, B. R. (2005). Gender differences in social support: A question of skill or responsiveness? *Journal of Personality and Social Psychology*, *88*(1), 79–90. https://doi.org/10.1037/0022-3514.88.1.79

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151–165. http://doi.org/10.3200/GENP.135.2.151-166

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92* (1–2), 91–112. https://doi.org/10.1016/j.jpubeco.2007.04.008

Raihani, N. J., & Bshary, R. (2015). Third-party punishers are rewarded, but third-party helpers even more so. *Frontiers in Behavioral Neuroscience*, *9*(39), 1–11. https://doi.org/10.3389/fnbeh.2015.00039

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 3677. https://doi.org/10.1038/ncomms4677

Romano, A., Balliet, D., Yamagishi, T., & Liu, J. H. (2017). Parochial trust and cooperation across 17 societies. *Proceedings of the National Academy of Sciences of the United States of America*, *14*(48), 12702–12707. https://doi.org/10.1073/pnas.1712921114

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin*, *132* (1), 98–131. https://doi.org/10.1037/0033-2909.132.1.98

Russell, A., Hart, C. H., Robinson, C. C., & Olsen, S. F. (2003). Children's sociable and aggressive behavior with peers: A comparison of the US and Australia, and contributions of temperament and parenting styles. *International Journal of Behavioral Development*, *27*(1), 74–86. https://doi.org/10.1080/01650250244000038

Schwartz, S. H., & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: Effects of gender equality. *Journal of Personality and Social Psychology*, *97*(1), 171–185. https://doi.org/10.1037/a0015546

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. https://doi.org/10.1006/jesp.1998.1373

Steele, J. R., & Ambady, N. (2006). "Math is Hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, *42*(4), 428–436. https://doi.org/10.1016/j.jesp.2005.06.003

Thielmann, I., Böhm, R., & Hilbig, B. E. (2015). Different games for different motives: Comment on Haesevoets, Folmer, and Van Hiel (2015). *European Journal of Personality*, *29*(4), 506–508. https://doi.org/10.1002/per.2007

Thielmann, I., Heck, D. W., & Hilbig, B. E. (2016). Anonymity and incentives: An investigation of techniques to reduce socially desirable responding in the Trust Game. *Judgment and Decision making*, *11*(5), 527–536. http://journal.sjdm.org/16/16613/jdm16613.pdf

Weng, H. Y., Fox, A. S., Hessenthaler, H. C., Stodola, D. E., & Davidson, R. J. (2015). The role of compassion in altruistic helping and punishment behavior. *PLoS ONE*, *10*(12), e0143794. https://doi.org/10.1371/journal.pone.0143794.g001

AUTHOR BIOGRAPHIES

Angela Dorrough is a postdoctoral researcher in social psychology at the University of Cologne. She worked as a research fellow at the MPI for Research on Collective Goods and graduated from the University of Göttingen. She combines methods of economics and psychology and is interested in social preferences and gender.

Maria Olsson is a doctoral researcher in social psychology at UiT The Arctic University of Norway. As part of her PhD, she spearheaded a large cross-national project on men's underrepresentation in communal roles. Her main research interests are gender roles, gender development, role models, and social inequality.

Laura Froehlich is a postdoctoral researcher in social psychology at the FernUniversität in Hagen, Germany. She graduated from the University of Konstanz in 2015. Her research interests are stereotypes about ethnicity and gender, immigrants' multiple social identities and cross-cultural approaches to social psychology.

Andreas Glöckner is professor of social psychology at the University of Cologne and senior research fellow at the Max Planck Institute for Research on Collective Goods, Bonn. He is editor in chief of the journal *Judgment and Decision Making* and past president of the European Association for Decision Making.

Sarah Martiny is a professor in social and community psychology at UiT The Arctic University of Norway. She graduated from the University of Jena and worked with Kay Deaux at New York University and as an assistant professor at the University of Konstanz. She is interested in migration, gender, and social inequality.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Dorrough AR, Olsson MIT, Froehlich L, Glöckner A, Martiny SE. Does she compensate the victim while he punishes the perpetrator? No gender differences in anonymous economic games across 11 nations. *J Behav Dec Making*. 2021;34:261–274. https://doi.org/10.1002/bdm.2208