

Acht, Martin; Rebaudo, Mara

**Article — Published Version**

## A gender gap in gender gaps: social norms and housework reporting

Empirical Economics

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Acht, Martin; Rebaudo, Mara (2025) : A gender gap in gender gaps: social norms and housework reporting, Empirical Economics, ISSN 1435-8921, Springer, Berlin, Heidelberg, Vol. 68, Iss. 6, pp. 2977-3029, <https://doi.org/10.1007/s00181-024-02710-z>

This Version is available at:

<https://hdl.handle.net/10419/323197>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# A gender gap in gender gaps: social norms and housework reporting

Martin Acht<sup>1,3</sup> · Mara Rebaudo<sup>1,2</sup> 

Received: 16 May 2024 / Accepted: 27 December 2024 / Published online: 6 February 2025  
© The Author(s) 2025

## Abstract

Gender differences in the amount of housework performed and the role of social norms in explaining these persistent gaps have received increasing attention from both policymakers and researchers in recent years. However, norms may not only affect the actual division of housework but also potentially influence the reporting behavior in surveys. We study how retrospective responses about time-use in face-to-face interviews are influenced by the gender of the interviewer. Our findings show that women tend to report significantly more hours of housework when interviewed by a woman rather than by a man. This effect is not observable for male respondents, resulting in an interviewer gender gap in the housework gender gap. Exploring the effect in relation to several norm-related characteristics indicates that social norms play an important role in the reporting of housework hours. Therefore, gender gap estimates based on face-to-face interviews should be interpreted with great caution.

**Keywords** Gender · Gender gap · Doing gender · Social norms · Interviewer effects

**JEL classification** D13 · J16 · D90

## 1 Introduction

The role of social norms as important drivers of human behavior has attracted the attention of researchers not least since the seminal work by Akerlof and Kranton (2000). In this field of research, a growing literature seeks to understand how social

---

The authors want to thank seminar participants at the EEA Congress, at the Annual Conference of the Verein für Socialpolitik, at the RGS Doctoral Conference, Marianna Schaubert and seminar participants in Sankt Augustin, Freiburg and Göttingen for helpful comments and suggestions.

---

✉ Mara Rebaudo  
mara.rebaudo@fit.fraunhofer.de

<sup>1</sup> Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany

<sup>2</sup> University of Freiburg, Freiburg im Breisgau, Germany

<sup>3</sup> Federal Ministry of Finance, Berlin, Germany

norms contribute to explaining persistent gender gaps in earnings, time-use and other outcomes. In their prominent contribution, Bertrand et al. (2015) demonstrate how couples engage in norm-conforming behavior by adjusting their labor market outcomes and time spent on housework to adhere to the traditional male breadwinner norm. In this study, we offer a new perspective on these findings and analyze whether and how social norms, which we define here as socially prescribed gender norms, affect the reporting of housework hours in surveys. Our approach exploits a contextual factor potentially related to the role of social norms in the response process: the gender of the interviewer. Specifically, we estimate respondent-fixed-effects models to study how the reporting of housework hours changes when the interviewer is female as opposed to male.

When studying time spent on housework, most evaluations rely on retrospective questionnaire responses rather than on diary-based data. There remains, however, discussion about how reliable such retrospective time-use data are. For example, Kan (2008) finds systematic differences between questionnaire and diary data on housework hours which vary by gender and other characteristics of the respondents. Several mechanisms to explain the gap between questionnaire-based and diary-based estimates regarding housework hours have been proposed in the literature. On the one hand, questionnaire-based estimates are likely to be biased due to an unclear definition of housework, recalling problems or double-counting when housework is performed simultaneously with other activities (Kan 2008; Press and Townsley 1998; Warner 1986). On the other hand, besides these more technical explanations, reporting housework hours in surveys is sometimes considered a gendered process. Press and Townsley (1998) argue that social desirability plays a central role in explaining differences between questionnaire and diary estimates, because some respondents may feel the pressure to report a level of housework that complies with normative gender roles. Our paper focuses on this aspect of housework reporting and is, to the best of our knowledge, the first to provide empirical evidence for whether and how norms affect the reporting of housework hours.

A growing body of literature points to the importance of social norms in explaining persistent gender differences in time spent on housework that cannot be rationalized within standard economic models. This relates to findings about how couples react to labor market events (Foster and Stratton 2018), to child birth or parental leave (Schober and Zoch 2019; Schober 2013), to entering retirement (Leopold and Skopek 2015), and also to the role of institutions in affecting gender differences (Cooke 2007; Lippmann et al. 2020). One of the most striking of these observations has been demonstrated by Bertrand et al. (2015): Standard economic models of decision-making within the family would predict that as a spouse's contribution to the couple's income increases, she should decrease her housework hours. However, Bertrand et al. (2015) show that women who outearn their partners and thereby violate the male breadwinner norm, actually *increase* their time spent on housework. This phenomenon has been interpreted in light of the "doing gender" theory first proposed by West and Zimmerman (1987). This theory posits that individuals "do gender" by actively conforming to societal expectations and norms associated with their perceived gender identity. When a traditional gender role is violated, for instance, when a wife outearns her husband,

individuals aim to compensate for this violation by reinforcing their gender identity in other ways, such as increasing their housework hours.

The presence of “doing gender” behavior has been established for several countries and contexts (e.g., Bertrand et al. 2015; Lippmann et al. 2020; Bittman et al. 2003; Foster and Stratton 2018). The finding that couples engage in norm-conforming behavior is not only relevant to the gendered division of housework hours, but also extends to women’s labor force participation and income (Bertrand et al. 2015). However, recent studies by Roth and Slotwinski (2020) and Murray-Close and Heggeness (2019) suggest that the observed norm-conforming behavior in this domain is the result of systematic misreporting of incomes. In light of this, while “doing gender” behavior regarding housework hours may indeed be present, it could be in terms of overreporting of housework hours rather than actual behavior (Sullivan 2011). Then, the aforementioned strategies to adhere to social norms may still be relevant, but probably in a different way than the literature currently suggests.

Analogous to the work by Roth and Slotwinski (2020), who compare official and reported income data for the same individuals, it would be ideal to compare retrospective responses about time-use to the “true” values to see if certain characteristics lead individuals to over- or underreport housework hours. As such data are not available and the credibility even of diary-based estimates remains questionable, we apply a different approach to study how social norms affect the reporting in interviews. Using data from the German Socio-Economic Panel, we study how retrospective responses about time-use in face-to-face interviews are influenced by the gender of the interviewer. If social norms play a role in the reporting of time-use, it seems likely that the characteristics of the interviewer, especially the interviewer’s gender, affect in which direction social norms influence the time-use estimates and how strongly. We focus on routine household chores such as washing, cooking and cleaning, as they are considered among the least enjoyable and also the most time consuming non-market activities (Sullivan 2013; Coltrane 2000), which points to the importance of social norms in explaining persistent gender differences in time spent on these tasks. For other activities the connection to social norms is less straightforward.

We find that female respondents report significantly more hours of housework when interviewed by a woman instead of a man. A similar gender-of-interviewer effect for male respondents is not observable. This in turn leads to an interviewer gender gap in the housework gender gap, because women and men respond differently to the interviewer’s gender. We interpret the findings as evidence for the fact that social desirability concerns – and hence social norms – play an important role in the reporting of housework hours. By comparing couples to singles, and also single-to-couple transitions over time, we disentangle the mechanisms behind socially desirable reporting, namely subconscious self-deception versus impression management. Stratifying the sample and interacting the interviewer’s gender with several norm-related characteristics, including respondents’ and interviewers’ cohorts, support the hypothesis that the observed effects are related to traditional social norms. We analyze differences by women’s relative income and employment status and thereby discuss how the interviewer’s gender is related to the finding by Bertrand et al. (2015) that couples “do gender” when women outearn their partners. Following the literature on differences between West and East Germans regarding social norms (Alesina and Fuchs-Schündeln 2007; Bauernschuster

and Rainer 2012; Beblo and Görges 2018), we find that gender-of-interviewer effects are less present for East Germans. Furthermore, we study differences in the gender-of-interviewer effect by source country gender equality of immigrants based on the literature on social norms and cultural heritage (Alesina et al. 2013; Blau et al. 2020; Giuliano 2020). In line with our hypothesis, the effects are smaller for immigrants from more gender-equal countries.

We provide the first causal evidence of how housework reports differ by the interviewer's gender, and how this gender-of-interviewer effect is related to social norms. Therefore, our paper contributes to both the strand of literature on interviewer effects and that on norms affecting the division of housework hours. We find differences in reporting behavior between female and male respondents, indicating that gender gap estimates based on face-to-face interviews should be interpreted with caution. We argue that the extent of gendered reporting is directly related to the internalization of traditional social norms. Thereby, we offer a new perspective on previous findings about the relationship between social norms and time spent on housework, as these may be influenced by reporting effects. While our approach focuses on reporting differences toward female versus male interviewers, our findings suggest that social norms may influence response behavior in even more ways than could be analyzed in this study.

The paper is structured as follows. Section 2 summarizes the existing evidence on gender-of-interviewer effects and provides a theoretical overview of the concept of social desirability bias to explain the findings. Section 3 describes the data and Sect. 4 presents the empirical strategy. Section 5 studies gender-of-interviewer effects on housework hours and relates the findings to several norm-related characteristics, including cohorts, employment status and income, differences between West and East Germany and the cultural heritage of immigrants. Section 6 provides robustness checks to address if any other factors other than the interviewer's gender are driving the results, before Sect. 7 concludes.

## 2 Background

Before summarizing the existing literature on gender-of-interviewer effects, we first provide a theoretical overview of the concept of *social desirability bias*, which provides the foundation for the interpretation of our results. The concept of socially desirable reporting suggests that respondents reflect on societal views and expectations regarding certain topics and adapt their responses accordingly (see Krumpal 2013, for an overview). Hence, instead of answering accurately and truthfully, some respondents distort their answers to conform to social norms and thereby maintain a socially favorable self-presentation. In fact, empirical studies have shown that respondents overreport socially desirable behavior, such as charity donations, voting, or seat belt use, and underreport socially undesirable behavior, such as smoking or illicit drug use (Tourangeau et al. 2000; Tourangeau and Yan 2007; Krumpal 2013).

The underlying processes and exact mechanisms of socially desirable reporting are still insufficiently understood (Tourangeau and Yan 2007; Holtgraves 2004). Scholars often distinguish between two types of norm-conforming reporting: Other-deception

or *impression management*, on the one hand, relates to the individuals' need for social approval by others, e.g., the interviewer (Paulhus 1991). It assumes that some respondents purposefully tailor their answers to avoid social disapproval or to create a positive image of themselves. *Self-deception*, on the other hand, refers to individuals' strategies to reduce cognitive dissonance caused by a divergence between social norms and reality (Krumpal 2013). As opposed to impression management, self-deception is mostly considered to be subconscious and automatic (Tourangeau and Yan 2007). While impression management behavior is directed at other subjects, e.g., interviewers, the main addressees of self-deceptive reporting behavior are the respondents themselves. Importantly, independent of the channel, socially desirable reporting will only occur when a deviation between the honest response and what is perceived to be desirable exists.

It has been shown that social desirability bias is especially likely in personal interviews and that interviewer characteristics, such as the interviewer's gender, can affect the likelihood of engaging in norm-conforming reporting behavior (Tourangeau et al. 2000; Tourangeau and Yan 2007). In terms of impression management, respondents may perceive female and male interviewers to have different opinions regarding gender-related questions and adjust their responses to gain social approval of the interviewer. However, even if respondents were not trying to impress their counterparts, the interviewer's gender can still affect the response in terms of self-deceptive behavior. Specifically, respondents may be more or less aware of their own deviation from social norms depending on their counterpart's gender.

Following this concept of socially desirable reporting, several studies analyze the role of gender-of-interviewer effects (see West and Blom 2017, for an overview). Most of the previous research has focused on normative views, for example about marriage (Liu and Stainback 2013) or women's issues and gender equality (Huddy et al. 1997). A common finding in the literature is that respondents either provide more feminist or liberal responses toward female interviewers or, the other way around, more traditional ones toward male interviewers (Kane and Macaulay 1993; Lueptow et al. 1990; Flores-Macias and Lawson 2008; Lipps and Lutz 2017).

It is important to note that previous studies on gender-of-interviewer effects almost exclusively rely on cross-sectional data and hence cannot establish causal effects. Furthermore, sample sizes are usually small and only a few interviewers conduct the surveys (Huddy et al. 1997). A major disadvantage of cross-sectional studies in this context is that interviewer effects are potentially confounded with area effects, as interviewers are typically allocated to specific geographic areas (Schnell and Kreuter 2005). To our knowledge, only two studies evaluating gender-of-interviewer effects related to gender issues address causality concerns by analyzing panel data, Zoch (2021) and Lipps and Lutz (2017). Zoch (2021) uses data from the National Educational Panel Study in Germany and analyzes how the interviewer's gender affects self-reported gender ideologies. In line with previous research on gender issues, she finds that respondents report less traditional attitudes toward female interviewers. According to this study, there are, however, no significant gender-of-interviewer effects for an attitudinal item asking whether it's the man's job to earn money and the woman's job to take care of the household and family. Housework hours or other time-use categories are not analyzed. The study by Lipps and Lutz (2017) is based on Swiss panel data

from telephone interviews and evaluates gender-of-interviewer effects on a variety of different topics. Some of the analyzed items are related to factual household tasks (e.g., “in our household, it is mostly me who does the cleaning”) and the authors also analyze hours of housework but do not find significant gender-of-interviewer effects for any of these items. Because it has a different focus, the study does not analyze how results may differ by subgroups or other norm-related characteristics. It is therefore possible that potential gender-of-interviewer effects are hidden, for example, because in our study we only find effects for couples and not for singles. Furthermore, the direct interaction with an interviewer (as opposed to a telephone interview) is likely heightening the salience of norms which we argue to drive the effects.

Our research question focuses on a factual survey question – time spent on housework – and hence differs from the majority of other studies about gender-of-interviewer effects that analyze attitudinal questions. Apart from Lipps and Lutz (2017) who do not find effects, there are no other studies evaluating gender-of-interviewer effects on housework hours. Importantly, housework hours are potentially very different from attitudinal questions about gender equality. In particular, the relationship between hours of housework and social norms is more subtle. While respondents may engage in impression management when answering attitudinal questions on gender equality for fear of offending the interviewer, this seems unlikely regarding the response about housework hours. Instead, we argue that the urge to fulfill a particular gender role is subconsciously affected by the interviewer’s presence. For example, respondents may become more or less aware of their own deviation from social norms depending on the interviewer’s gender. In line with self-deceptive reporting behavior this perception may subsequently influence the thought process of finding and formulating an answer, given that most people are uncertain about their exact housework hours. Our empirical analysis of different subgroups helps to disentangle the possible mechanisms for socially desirable reporting and supports this hypothesis of self-deceptive behavior.

### 3 Data

We use data from the German Socio-Economic Panel (SOEP), a longitudinal survey of about 15,000 private households in Germany with annual interviews since 1984 (Goebel et al. 2019).<sup>1</sup> The majority of interviews in the SOEP are face-to-face interviews, which allows us to study the effect of the interviewer’s gender on individual responses. Notably, SOEP interviews are highly standardized. Interviewers are employed by the field institute Kantar, which prioritizes interviewer monitoring through ISO-certified processes that are regularly audited (Bohlender et al. 2020). For instance, at least 10% of annual interviews are checked, and each interviewer is monitored at least once a year. Interviewers are trained to ensure high data quality. They participate in annual training sessions conducted by the SOEP team at Kantar, along with representatives from DIW Berlin, where SOEP is based. These sessions prepare contact interviewers, who then train interviewers in the regions for which they are responsible. Contents of these trainings are scope, timing, and procedures of the

<sup>1</sup> Socio-Economic Panel (SOEP), data for years 1984-2018, version 35, 2020.



interviews, as well as any special features specific to each survey year (Goebel et al. 2019; Bohlender et al. 2020).

Our main outcome of interest is the time spent on housework, which is measured using the following question (see the entire question in Fig. 1 in the Appendix): “*What is a typical weekday like for you? How many hours per normal workday do you spend on housework (washing, cooking, cleaning)?*”. This definition of housework aligns with the typical usage in the literature (e.g., Lippmann et al. 2020), specifically excluding other non-market activities like childcare or repair work and gardening, which are listed as separate activities in the same interview question. Routine housework is considered by both women and men as one of the least enjoyable activities (Gershuny 2013; Sullivan 2013), making it an interesting subject to study in the context of within-couple differences. Tasks like repair work and gardening, unlike routine housework, have been found to be more time flexible and more enjoyable (Coltrane 2000). As argued by Sullivan (2010), the norm concerning childcare has shifted since the 1970s, with time spent with children increasingly being viewed as leisure, particularly among more educated parents. Given that routine housework is most directly related to social norms, we restrict our focus on housework hours for the main specification. However, we also provide robustness checks for the other activities listed in the time-use question.

Our analysis covers all years from 1985 to 2018, as information about the interviewers is available from 1985 onwards. We follow Lippmann et al. (2020) and Bertrand et al. (2015), and focus on working-age individuals between 18 and 65 years of age, since we can analyze the role of employment status and (relative) income for this group. The main sample consists of cohabiting couples (married or not married), although additional results for singles will be discussed in section 5. Importantly, in the SOEP data, all adults in the household answer the personal questionnaire, including the time-use reports, themselves. Thereby, within-couple gaps in housework reports can be analyzed. We exclude the few observations of partners who were interviewed by different interviewers.<sup>2</sup> We furthermore exclude more-generation-households and other non-typical household arrangements. Because the interviewer’s gender may differently affect homosexual individuals, we also exclude them from the analysis.<sup>3</sup> Our final sample consists of 18,965 couples that are observed for an average of around six years. Table 1 presents descriptive statistics of the main sample. While women spend on average about three hours per weekday on housework duties, the average for men is 46 min per day.<sup>4</sup> Of all individuals considered (employed and unemployed), the average working time of women is nearly half that of men, and incomes are consistently lower. Men in our sample are a bit older and on average have a slightly higher

<sup>2</sup> In the final sample this concerns only 68 observations. Couples are not excluded if partners are interviewed on different dates. In our sample 88% of couples are interviewed on the same day and around 95% within seven days after the first partner’s interview. Results are almost identical when excluding couples who were interviewed on different dates, see Table 11 in the Appendix.

<sup>3</sup> In our sample, there are only 112 different same-sex couples.

<sup>4</sup> In our sample women’s maximum housework time is 24 h and men’s is 16 h per day. When we assume 8 h of sleep, housework estimates of more than 16 h are likely overstated. However, only 32 observations report female housework hours to be larger than 16. Our results are robust when observations reporting high levels of housework, such as more than 16 h or even more than 6 h, are excluded, see Table 12 in the Appendix.



**Table 1** Descriptive statistics

	Mean	Standard deviation	Min	Max
Woman's housework time	3.00	1.95	0	24
Man's housework time	0.76	0.93	0	16
Housework woman - housework man	2.24	2.23	-15	24
Paid work time woman	19.47	18.08	0	80
Paid work time man	36.69	18.69	0	80
Income woman	775.25	908.06	0	36184
Income man	2054.58	1725.40	0	98038
Woman's age	42.58	10.72	18	65
Man's age	45.33	10.77	18	65
Education level woman	4.92	2.33	0	9
Education level man	5.01	2.43	0	9
Married	0.88	0.33	0	1
No. of children age <6	0.30	0.59	0	6
No. of children age 6-12	0.43	0.73	0	5
No. of children age 13-16	0.23	0.50	0	4
Income HH	3447.94	2062.53	43	210084
East Germany	0.21	0.41	0	1
Observations	110,544			

The table reports summary statistics of respondents' characteristics based on the main sample (cohabiting couples of age 18 to 65). Housework time is hours per day. Paid work time is hours worked per week including overtime hours. Income is net monthly income and household income is net monthly income of all household members including non-labor income. Income is corrected for inflation with base year 2016. Education levels derive from the CASMIN classification. Married is a dummy variable equal to 1 for married couples. No. of children is the number of children living in the household for each of the age groups. East Germany is a dummy variable equal to 1 when the couple lives in East Germany at the time of the survey

education level. 88% of the couples observed are married and around 21% live in East Germany at the time of the survey.

In total, our sample consists of 1,841 different interviewers. Around 57% of all interviews are performed by male interviewers. Female and male interviewers differ in some characteristics that may affect interview responses, as shown in Table 2. Male interviewers on average have slightly less experience with SOEP, are older, more often married, more likely to have a higher education level, more often born in Germany, and more often of German mother tongue. Unfortunately, most questions about additional interviewer characteristics, except the interviewer's age and experience, are only asked in specific years, so that including them as controls in the regression analyses results in a more selective sample. Therefore, we only include the interviewer's gender and age for the main analysis and provide specifications with the additional characteristics as robustness checks in section 6.<sup>5</sup> Apart from the observed differences between

<sup>5</sup> Interviewer's experience does not affect housework reports and the gender-of-interviewer effect is unaffected when additionally controlling for experience, see Table 13 in the Appendix. We do not include it as

**Table 2** Interviewer characteristics by interviewer gender

	Mean male interviewer	Mean female interviewer	Diff	SE diff	Observations
Age	60.31	57.85	2.457***	0.069	110544
Years experience with SOEP	9.97	10.44	−0.468***	0.044	110544
Married	0.84	0.68	0.160***	0.003	72278
Higher education	0.40	0.30	0.094***	0.004	70440
Born in Germany	0.97	0.93	0.040***	0.002	70201
Mother tongue German	0.99	0.94	0.049***	0.001	62154

The table reports mean comparisons of interviewers' characteristics by interviewer gender based on the main sample (cohabiting couples of age 18 to 65). Only the interviewer's age and experience is observed for all interviewers, which is why the sample size differs from the main sample for all other characteristics. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

female and male interviewers, interviewers in our sample also differ from respondents, especially with regard to their age. On average, interviewers are more than ten years older than respondents. We discuss how this may affect the results in section 5.3 when analyzing differences by respondent and interviewer cohorts.

## 4 Empirical strategy

We are interested in the effect of the interviewer's gender on the reported amount of housework hours. Hence, the latter will be our main dependent variable and the interviewer's gender is included as an explanatory variable. By running individual fixed-effects models we estimate how housework reports differ when respondents are interviewed by a female versus a male interviewer and account for unobserved differences between respondents. In our specification, we follow Lippmann et al. (2020) and Bertrand et al. (2015) and control for several factors that influence non-market work. Formally, we estimate the following equation for the sample of women, men, or couples  $i$  and time  $t$ ,

$$\begin{aligned}
 \text{HouseworkHours}_{it} = & \alpha + \beta \text{FemaleInterviewer}_{it} \\
 & + \gamma \text{InterviewerCharacteristics}_{it} \\
 & + \delta X_{it} + \mu_i + \lambda_t + \epsilon_{it}
 \end{aligned} \quad (1)$$

with *FemaleInterviewer* being a dummy variable coded one if the interviewer is female and zero if the interviewer is male. *InterviewerCharacteristics* in the main specification only includes the interviewer's age as explained above. Respondent controls included in  $X$  are paid work time (including overtime hours), net income, age, age squared, and education of both partners in the couple, as well as their marital status, the

control variable in the main specification to avoid potential collinearity issues, as interviewer experience increases by 1 with every survey year in which the interviewer does not change, similarly to interviewer and respondent age.

number of children of different age groups and household income. Additionally, we control for the length of the interview of both partners, and for the interview mode.<sup>6</sup> In all regressions, we include federal state, year ( $\lambda_t$ ), and individual fixed-effects ( $\mu_i$ ).<sup>7</sup>

We argue that interviewer gender can be interpreted as exogenous variable, at least after controlling for the relevant variables that we include in our regression analyses. To support this, Table 14 in the Appendix displays how several characteristics of the respondents differ by the interviewer's gender. We find statistically significant differences only in a few of the considered characteristics, such as net income, working hours, age, and education, and the differences are economically small. Other statistically significant yet small differences regard the length of the interviews (which is included as control variable), and the timing of the interview which we discuss further in section 6. Importantly, the empirical approach of estimating respondent fixed-effects regressions does not require the allocation of interviewers to be completely random. For example, interviewers are typically assigned to specific geographic areas, so that the interviewers in one region may differ from interviewers in another region. This may lead to small differences in average respondent characteristics by interviewer gender if the gendered allocation of interviewers is not completely random. However, the respondent-fixed-effects models do not exploit variation across different respondents (from different geographic regions), but instead, they focus on the effect of interviewer changes for the same individuals. The gender-of-interviewer coefficient is only biased if a change in interviewer gender coincides with a simultaneous change in an unobserved variable that also affects housework hours. We argue that this is unlikely, and address any potential concerns in section 6. For these reasons, we argue that a significant coefficient for *FemaleInterviewer* can be interpreted as the causal effect of the interviewer's gender on housework reports.

## 5 Results

### 5.1 Gender-of-interviewer effects on housework reports

We begin by estimating equation (1) for the sample of cohabiting couples. Results are presented in Table 3. The dependent variable is individual daily housework hours of women and men for columns 1 and 2 and the housework gap within the couple, measured as *HouseworkHours Woman* – *HouseworkHours Man*, in column 3. Of major interest to our research question is the effect of the interviewer's gender on the housework report: We find that for the subsample of women the coefficient of the

<sup>6</sup> The presence of other people during the interview (e.g., the respondent's partner) may also be relevant. Unfortunately, this information is not available in the SOEP data for almost all of the studied years and hence can not be analyzed in this study.

<sup>7</sup> Some of the explanatory variables in the regression analyses could potentially be influenced by the interviewer's gender itself, e.g., income, which has been shown to be biased by reporting effects (Roth and Slotwinski 2020). We ran regression analyses to test the effect of the interviewer's gender on income and did not find significant effects. Results are available on request. Furthermore, variables like income and hours of work are likely chosen jointly with housework hours. We are, however, not trying to estimate a collective model of paid work and household labor supply but instead only focus on the separate effect of interviewer gender on housework hours.

**Table 3** Housework time and interviewer gender

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.079*** (0.02)	0.014 (0.01)	0.065*** (0.02)
<i>Respondent characteristics</i>			
Paid work time woman	−0.032*** (0.00)	0.005*** (0.00)	−0.037*** (0.00)
Paid work time man	0.004*** (0.00)	−0.015*** (0.00)	0.020*** (0.00)
Income woman	−0.000*** (0.00)	0.000*** (0.00)	−0.000*** (0.00)
Income man	0.000*** (0.00)	−0.000*** (0.00)	0.000*** (0.00)
Log. HH income	0.012 (0.03)	−0.022 (0.02)	0.034 (0.03)
Married	0.068** (0.03)	−0.058*** (0.02)	0.127*** (0.03)
No. of children age <6	0.363*** (0.02)	0.045*** (0.01)	0.317*** (0.02)
No. of children age 6-12	0.245*** (0.01)	0.022*** (0.01)	0.223*** (0.02)
No. of children age 13-16	0.147*** (0.01)	0.020** (0.01)	0.127*** (0.02)
<i>Interviewer characteristics</i>			
Age	−0.000 (0.00)	0.000 (0.00)	−0.000 (0.00)
Interview length woman	0.000 (0.00)	0.000 (0.00)	−0.000 (0.00)
Interview length man	0.002*** (0.00)	0.000 (0.00)	0.001** (0.00)
Adjusted $R^2$	0.167	0.097	0.169
Individuals	18965	18965	18965
Observations	110544	110544	110544

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHours Woman − HouseworkHours Man*). The table reports the results of individual (or couple) fixed effects estimations. Additional controls include age, age squared, education categories for women and men, respectively, interview mode, as well as year and state fixed effects. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

interviewer's gender is statistically significant and positive. Women report on average 5 min more housework time per day when they are interviewed by a woman instead of a man. Compared to an average housework time of 3 h per day (see Table 1) this makes up roughly 3%. For men, we do not find an effect. This result is confirmed by the gender-of-interviewer coefficient being significant and positive, albeit smaller, for the within-couple difference in housework hours. Hence, when a couple is interviewed by a woman instead of a man the gap between female and male housework hours increases by about 4 min per day (again roughly 3% compared to the sample mean).<sup>8</sup>

Note that in the SOEP data time-use is generally measured in full hours, so that small over- or underreports to fulfill social desirability are not possible. This creates a larger obstacle for individuals to adjust their reports. Furthermore, as outlined in section 2, socially desirable reporting not only depends on an individual's inclination to engage in norm-conforming behavior, but also on the existence of a deviation between the honest response and what is perceived as the social norm. Therefore, not everyone will feel compelled to adjust their report in response to the interviewer's gender. The fact that we nevertheless find statistically significant effects suggests that some individuals do indeed adjust their responses by full hours when confronted with an interviewer of a different gender. It seems likely that the observed effect underestimates the effect that may occur if respondents could adjust their reports by smaller increments. While smaller adjustments in the housework reports (compared to adjustments by full hours) may decrease the observed effect at an individual level, the more relevant factor is likely the number of respondents adjusting their reports, which is expected to be much larger when the barrier for over- or underreporting is smaller.

We interpret the observed gender-of-interviewer effect as female respondents reporting *more* housework hours to *female* interviewers. While it is possible that the effect goes in the other direction by female respondents reporting *fewer* housework hours to *male* interviewers, we argue that the results and additional explanations that follow are more in line with the former interpretation.

The coefficients for respondents' control variables in Table 3 are in line with the ones usually reported in the literature: For women and men more hours of paid work decrease hours spent on housework, while their partner's paid work time increase individual housework hours. The results for the within-couple gap suggest that the effect on female housework hours tends to be stronger, so that increases in women's paid work time decrease the couple difference, while increases in male paid work time increase it. Similar effects are observed for the role of women's and men's income. Marriage is associated with women doing more and men doing less housework, increasing the within-couple gap. The number of children increases housework hours for both women and men with the effect being larger when there are younger children. Again, the positive effect is larger for women, resulting in a positive effect of the number of children on the gender gap.

The main results are robust when using different definitions for housework, for example the sum of household chores (housework, repairs/gardening, and errands), or when analyzing total non-market work (additionally including childcare and elderly

<sup>8</sup> The gender-of-interviewer coefficient is very similar in the cross section and for regressions excluding arguably endogenous controls, such as hours of work and income, see Tables 15 and 16 in the Appendix. Additionally, the effect is similar when only including dual-earner or married couples.

care), see Table 17 in the Appendix. The magnitude and significance level of the gender-of-interviewer effect changes, however. To gain more insight into how the interviewer's gender impacts the different activities covered by the same interview question about time-use, we estimate separate regression models for each activity in Table 18 in the Appendix. For most of the analyzed activities, we do not find any gender-of-interviewer effects neither for female nor for male respondents. This alleviates potential concerns that female interviewers in general behave differently than male interviewers, for example by discussing more clarifying questions with female respondents. We do find a strong and positive effect for women reporting more childcare hours toward female interviewers, which is in line with the effect on housework hours. For elderly care, we do not find significant effects, which may, however, be driven by a lack of variation in time spent on elderly care. Only a small subsample (less than 5% of observations) reports any positive hours for elderly care, and it also occurs only for limited periods. In addition to the positive impact on childcare, there is a marginally significant and much smaller effect for women reporting more hours spent on "repairs on and around the house, car repairs, garden work" toward female interviewers. This is surprising at first glance, as at least repair work would typically not be considered a female task. Gardening, however, is shared more equally between women and men compared to other activities, (e.g., Jessen et al. (2024)), making it less clear which effects to expect. It does not seem unlikely that at least for gardening, social desirability concerns may play some role. Unfortunately, the data do not allow to distinguish between repairs and gardening, which may differ with respect to their implied social roles. Interestingly, even for repairs and gardening, we do not find gender-of-interviewer effects for male respondents.

Furthermore, we find that women and men report more time spent on "leisure activities" toward female interviewers (or less leisure time toward male interviewers). While it is beyond the scope of this paper to extensively discuss potential mechanisms, a few points can be made: While less prominent in the literature than housework activities, there is a gender gap for leisure time with men spending more time on leisure than women (e.g., Beblo and Robledo 2008). Hence, social desirability concerns when responding to this question are not unlikely. For example, women may feel the urge to "do it all" by being a good housewife who still has time for hobbies. Another possible mechanism might be that respondents report less leisure time to male interviewers, for example, men when interviewed by men may like to be perceived as career-focused not "wasting" time on leisure. No effects are found for "exercise," which has, however, only been included as separate activity since 2013, which may explain the absence of effects. For the remainder of the paper, we proceed with housework hours as the dependent variable of interest.

## 5.2 Subconscious self-deception versus other channels

Our main result suggests that women report more housework hours to female interviewers, while we do not find any effects for male respondents. Given the discussion in section 2, this result seems to be surprising at first. Previous results on gender-of-interviewer effects concerning normative questions tend to find that women report

more feminist or liberal views to female interviewers. However, as we have argued, factual items such as hours of housework are potentially very different from attitudinal items and more subtly related to social norms. Consequently, it appears unlikely that respondents intentionally tailor their answer out of fear of offending the interviewer. Our finding that women tend to report more housework hours when interviewed by a woman rather than a man suggests that women feel a stronger urge to fulfill traditional domestic roles when talking to other women. This could be a result of female respondents expecting other women to have more traditional views compared to men. However, this seems unlikely given the findings from gender-of-interviewer effects on normative view questions (which imply that women expect female interviewers to have less traditional views than male interviewers) and also contrasts with analyses of attitudinal questions about gender roles, which show that women, on average, hold less traditional views than men (Bauernschuster and Rainer 2012). An alternate explanation for the observed effect is that women internalize societal expectations more strongly when interacting with a female interviewer as opposed to a male one. The likely mechanism here is that most respondents are unaware of how much time they actually spend on housework and must therefore engage in some thought process to find an answer. The social roles individuals aim to conform to influence this thought process, which in turn may be reinforced by the presence of a female interviewer. Thus, we hypothesize that the subconscious urge to fulfill a certain gender role is affected by the presence of the interviewer, subsequently influencing the response in line with self-deceptive reporting behavior.

That we do not find reporting effects for male respondents suggests that men do not feel compelled to conform to specific gender roles by adjusting their housework reports. This could either indicate that men do not experience the same pressure as women to adhere to societal expectations regarding their participation in housework duties, or, and potentially related, they do not believe there to be any such expectations. The idea that norms concerning involvement in housework hours affect men less than women is supported by the findings from Lippmann et al. (2020) and Bertrand et al. (2015), which show that male housework hours are less responsive to violations of the male breadwinner norm. It is possible that men might engage in norm-conforming behavior in other areas where they perceive stronger societal expectations about their presumed social roles.

We interpret these results as evidence that social desirability concerns and hence social norms affect the reporting of housework hours. To further support this interpretation, we compare the main results to single women and men for whom social norms regarding housework hours are presumably less relevant than they are for couples (Geist 2009). Results are presented in Table 4. As can be seen from the first two columns, neither single women nor men report differently toward female interviewers. To study the link between singles and couples, in columns 3 and 4 we control for single-to-couple transitions.<sup>9</sup> Thus, we include all singles and couples fulfilling the other sample criteria and additionally control for a dummy variable *Couple* coded one if the individual is in a couple relationship and zero otherwise. As can be seen from

<sup>9</sup> The SOEP data attempts at following individuals after separations, so that both single-to-couple transitions and couple-to-single transitions are captured.



the estimated coefficients, being in a couple relationship as opposed to being single, for women results in more housework hours, while for men it results in less. More interestingly for our purpose, the interaction effect between the *Couple* dummy and the interviewer's gender for women is statistically significant and positive. Hence, the established result that the gender-of-interviewer effect is present for women in couple relationships but not for singles can also be found when studying transitions within individuals over time. As expected from the previous results there is no statistically significant interaction effect for men.

The comparison of couples and singles helps to further disentangle the precise mechanisms behind socially desirable reporting, namely self-deception versus impression management. Above, we have argued that it seems unlikely for respondents to fear offending the interviewer, as housework hours are different from attitudinal items on gender equality. There are two additional explanations consistent with impression management as to why respondents may adjust their housework reports based on the gender of the interviewer. Firstly, findings on experimenter-gender-effects from other sciences suggest that there might be a general desire to impress one's counterpart, especially in opposite-sex relationships (for heterosexual individuals) (Chapman et al. 2018). Individuals want to be perceived positively by the opposite sex, which can lead to self-affirming feedback or even the possibility of a romantic relationship. In relation to our results, this mechanism could indeed play a role if women were reporting more housework toward male interviewers, but it seems implausible that reporting *less* toward male interviewers would improve women's perceived attractiveness as potential partners. Furthermore, if such a channel were in place, it should also be apparent for single women, since attractiveness on the marriage market is likely even more relevant for them. Hence, both the direction of the effect and the non-existence of effects for single women contradict such a channel.

Secondly, and also in line with impression management theory, respondents may have a general desire to not being considered as unclean by other individuals when their own housework hours are perceived too low. If women expect higher cleanliness standards from other women as opposed to men, this channel could be in line with higher housework reports toward female interviewers. However, if those concerns were indeed driving the results, again one would expect similar effects for single women. As we do not find those, general expectations regarding cleanliness standards also do not seem to explain our results. As a consequence, subconscious self-deception instead of impression management seems to be the more relevant channel to explain the observed effect.

By comparing couples versus singles, we compare gender-of-interviewer effects for subgroups which also differ with respect to their total time spent on housework. Therefore, we need to address the issue of absolute versus relative adjustments in the housework report. For instance, if the extent of gendered reporting depended on the baseline level of housework, the absolute gender-of-interviewer effect might only be small and insignificant for singles because they do less housework than women in couple relationships. Therefore, in Table 19 in the Appendix we analyze the gender-of-interviewer effect in relative terms by using the logarithm of housework hours as the dependent variable. The results are qualitatively and quantitatively in line with the main

**Table 4** Housework time and interviewer gender - singles

	Singles		Single-couple	
	Women (1)	Men (2)	Women (3)	Men (4)
Female interviewer	-0.046 (0.03)	0.018 (0.03)	-0.005 (0.03)	-0.005 (0.02)
Couple			0.190*** (0.03)	-0.396*** (0.02)
Female interviewer * couple			0.073** (0.03)	0.023 (0.03)
Adjusted $R^2$	0.101	0.054	0.171	0.088
Individuals	7564	4783	23668	21486
Observations	33912	20334	144456	130878

Dependent variable is female or male daily housework hours. The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

specification. Importantly, even in relative terms there are no gender-of-interviewer effects for men or singles.

To summarize, women in couple relationships report more housework hours toward female interviewers which we interpret as evidence for social norms affecting housework reports. Following the theory on social desirability, we expect gender-of-interviewer effects to be larger when the deviation between what is perceived to be desirable and reality is larger. Moreover, gender-of-interviewer effects may depend on the internalization of traditional social norms. These hypotheses will be examined in the subsequent sections.

### 5.3 Gendered reporting and norm-related characteristics

If the reporting of housework hours is driven by the desire to fulfill a certain role, the size of the effect may differ by different norm-related characteristics. Studies show that social norms can evolve over time (Fernández 2013) so that later-born generations possibly hold less traditional views about the role of women and men. With regard to our research question, it is therefore interesting to analyze whether the gender-of-interviewer effect differs by respondent and interviewer cohorts.

Table 5 presents the results. We split respondents up into two groups by the threshold of 1960 as year of birth, which is roughly the sample mean. As columns 1 and 2 show, the gender-of-interviewer coefficient indeed appears to be smaller for the later born cohort of women, despite still being marginally significant. In column 3 we analyze how the effect differs by cohorts of interviewers. With respondent fixed-effects models we can observe variation in the interviewer's cohort over time (but not in the respondents' cohorts), so we split interviewers up into three instead of two groups. In

**Table 5** Women's housework time - respondent and interviewer cohorts

	Young cohort (1)	Old cohort (2)	All (3)	Young cohort (4)	Old cohort (5)
Female interviewer	0.049* (0.03)	0.117*** (0.04)	0.103*** (0.03)	0.078* (0.04)	0.121*** (0.05)
[Reference: Interviewer born < 1945]					
Interviewer born 1945-1959					
Interviewer born >= 1960			-0.001 (0.05)	0.044 (0.05)	-0.063 (0.08)
Female interviewer			-0.017 (0.08)	-0.022 (0.09)	0.040 (0.14)
* Interviewer born 1945-1959			-0.015 (0.05)	-0.026 (0.07)	0.029 (0.08)
Female interviewer			-0.168** (0.08)	-0.150* (0.08)	-0.163 (0.16)
* Interviewer born >= 1960					
Adjusted $R^2$	0.156	0.184	0.167	0.156	0.184
Individuals	12027	6938	18965	12027	6938
Observations	63757	46787	110544	63757	46787

Dependent variable is female daily housework hours. The sample is restricted to cohabiting couples (age 18 to 65). Columns are based on respondent cohorts: Young cohort is defined as at least one person in the couple being born 1960 or later. The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

addition to the 1960-threshold that is comparable to respondents' cohorts, we separately account for interviewers born before 1945 and interviewers born between 1945 and 1959. The gender-of-interviewer coefficient persists toward interviewers of the two older cohorts, which represent more than 90% of interviewers in our sample. However, there is a reverse effect for the youngest cohort of interviewers, resulting in an insignificant gender-of-interviewer coefficient for respondents being interviewed by these interviewers. When we compare the effects of the interviewer cohorts by the two respondent cohorts (columns 4 and 5), it appears that especially respondents from the younger cohort engage less in gendered reporting when interviewed by the youngest generation of interviewers, however, there may also be more transitions toward young interviewers in this group. The resulting gender-of-interviewer coefficient for the young respondent cohort toward young interviewers is not statistically different from zero. For male respondents, we still do not find any gender-of-interviewer effects when separately studying results by cohorts (see Table 20 in the Appendix). Summarizing, there is some evidence that gender-of-interviewer reporting is less present when respondents and more importantly interviewers are from younger cohorts and hence may hold less traditional views. This supports our hypothesis that gendered reporting is indeed related to social norms. We do not find that the interviewer's age (independent of the birth cohort) is related to the gender-of-interviewer effect (see Table 21 in the Appendix).

Furthermore, we analyze how results differ by education levels and by the presence of young children, as both may be related to the role respondents wish to fulfill. For example, mothers of small children may have a stronger urge to be perceived as "good" mothers by reporting more housework hours. Highly educated women, in contrast, may feel less pressure to conform to traditional roles. Interestingly, we find that the gender-of-interviewer effect does not differ between respondents of low or high education and also not by the presence of children in the household (see Tables 22 and 23 in the Appendix). This further corroborates that gender-of-interviewer effects not only depend on respondents' perceived social norms, but also on the deviation between the honest response and what is perceived to be desirable. By children and education status, such a deviation does not seem to differ much, despite potential differences regarding perceived social norms.

## 5.4 Gendered reporting and specialization in paid work

If the reporting of housework hours is driven by the desire to fulfill a certain role, the size of the effect may differ by women's employment status and income. We suppose that a woman's employment status (in addition to her working hours) may affect her belief about which social role to fulfill: For example, full-time employed women – compared to other women – might feel less pressure to conform to the housewife role and hence feel less urge to (subconsciously) adjust housework reports. Then, the gender-of-interviewer effect that potentially captures the need to fulfill societal expectations, would be smaller for full-time employed women. Accordingly, a woman's contribution to the couple's income might also affect her belief about societal expectations regarding housework. In line with the previous hypothesis, it could be that

women with a higher contribution to household income feel less need to adjust housework reports to fulfill the housewife role because they are identifying themselves more as working women. In contrast, the “doing gender” literature would suggest that a woman earning more than her partner feels the need to conform to her traditional role by either doing more housework or, potentially, reporting more housework hours (Lippmann et al. 2020; Bertrand et al. 2015). According to this idea, women who out-earn their male partners might show a *stronger* gender-of-interviewer effect because of the desire to conform to their social role.

To test these hypotheses, we add women’s employment status and relative income as controls to the regression analyses. Results are presented in Table 6. For these regressions, we restrict the sample to dual-earner couples as the interpretation of differences in relative income and employment status is only meaningful for this group. In contrast to relative income, we define the full-time dummy in absolute terms for females (not in relative terms within the couple) because in our sample of dual-earner couples, 95% of men work full-time but only 43% of women do. Whether the woman is working full-time and her relative income are both included in the regressions as interaction terms with the interviewer’s gender and also used for subgroup analyses. Consistent with our first hypothesis, running separate regressions for full-time and not-full-time employed (part-time/irregularly employed) women (columns 1 and 2) shows that the gender-of-interviewer effect of the latter is almost twice as large as of the former. Hence, it seems to be the case that full-time employed women feel less need to adjust housework reports. Including women’s employment status as an interaction effect with the interviewer’s gender (column 3) suggests that full-time employment generally reduces housework hours, but the relationship between full-time employment and the interviewer’s gender is not statistically significant.

To study the role of relative income we define two variables, following Bertrand et al. (2015). Firstly, the variable *RelativeIncomeWoman* measures the woman’s share in the couple’s income, defined as  $IncomeWoman / (IncomeWoman + IncomeMan)$ . Secondly, we define a dummy variable *WomanEarnsMore* coded one when the woman earns strictly more than her partner, and zero otherwise. When separately studying couples where the man earns more (or incomes are equal) and couples where the woman earns more (columns 4 and 5), we see that the gender-of-interviewer effect is smaller and not statistically significant in the latter group. This supports the first hypothesis that women with a higher contribution to the couple’s income feel less need to adjust their housework reports to fulfill a certain role. Furthermore, we run two additional regressions on the whole sample: In column 6, we control for women’s relative income and additionally include an interaction effect between the interviewer’s gender and a dummy variable indicating whether the woman earns more than the man.<sup>10</sup> The relationships shown by Bertrand et al. (2015) and Lippmann et al. (2020) are also present when additionally controlling for the interviewer’s gender: Larger relative income decreases female housework hours, but there is a statistically significant

<sup>10</sup> This specification is similar to the ones by Bertrand et al. (2015) and Lippmann et al. (2020) as they control for the woman’s income share and additionally for a dummy variable indicating whether the woman earns more than the man. They do so in order to capture only the effect associated with the violation of the social norm that “a man should earn more than his wife” irrespective of specialization effects related to relative income.

**Table 6** Women's housework time and interactions - dual-earner couples

	Full-time (1)	Employment Status Not full-time (2)	All (3)	Male higher or equal (4)	Relative Income Female higher (5)	All (6)	All (7)
Female interviewer	0.058** (0.03)	0.107*** (0.03)	0.096*** (0.03)	0.088*** (0.03)	0.042 (0.04)	0.079*** (0.02)	0.130*** (0.04)
Female interviewer * Woman full-time			-0.047 (0.03)				
Woman full-time			-0.121*** (0.03)				
Female interviewer * Woman earns more						-0.033 (0.03)	
Female interviewer * Relative income woman							-0.167* (0.10)
Woman Earns More						0.127*** (0.02)	0.113*** (0.02)
Relative income woman						-1.148*** (0.10)	-1.077*** (0.11)
Adjusted $R^2$	0.045	0.088	0.099	0.101	0.048	0.100	0.100
Individuals	7032	8890	13145	11861	3349	13145	13145
Observations	26695	34962	61657	51779	9878	61657	61657

Dependent variable is female daily housework hours. The sample is restricted to dual-earner couples (age 18 to 65). The table reports the results of individual fixed effects estimations. Respondents and interviewer controls as well as year and state fixed effects are included. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

and positive coefficient for the *WomanEarnsMore* dummy, which represents the so-called doing gender effect. The interaction effect of the interviewer's gender with the *WomanEarnsMore* dummy in this regression is not statistically significant. When the interviewer's gender is interacted with relative income (instead of the dummy variable) (column 7), we observe a negative and marginally significant interaction effect. Thus, the higher the woman's contribution to the couple's income, the less they engage in gendered reporting. In both specifications, the overall gender-of-interviewer coefficient remains positive and statistically significant.

These results combined with the previous ones suggest that women who are relatively more attached to paid work, feel *less* need to report more housework hours toward female interviewers. This is in line with the expectation that gender-of-interviewer effects will be larger when the deviation between what is expected to be desirable and reality is larger. While we focus on dual-earner couples for these analyses, Table 24 in the Appendix also evaluates whether being employed (vs. not being employed) affects the results. As shown, the gender-of-interviewer effect does not differ by employment status at the extensive margin. As unemployed women typically do more housework than employed women, it is likely that they do not perceive a stronger deviation between their actual hours and their expected hours, compared to working women.

We do not find evidence for the second hypothesis that women earning more than their partners react more strongly to the interviewer's gender. Note, however, that these results neither confirm nor contradict the general idea of "doing gender" reporting, where women violating traditional social norms in the market sphere (by outearning their partners) report more housework hours to restore their gender identity. It could, for example, still be the case that these women report more housework hours, but do so similarly when confronted with female and male interviewers.

Besides employment status and relative income, regional differences present another category of norm-proxies to test their relation to gender-of-interviewer effects: Following Alesina and Fuchs-Schündeln (2007), several studies exploit the natural experiment constituted by the 41-year division of Germany to study differences in social norms between Western and Eastern regions. Among studies evaluating gender attitudes and norms, results show that East Germans tend to hold more egalitarian gender-role attitudes than West Germans (Bauernschuster and Rainer 2012; Beblo and Görges 2018). Lippmann et al. (2020) analyze how the results presented by Bertrand et al. (2015) differ between West and East Germany and find substantial differences. Specifically, they show that "doing gender" behavior, i. e. the urge to compensate for violations of the "male breadwinner" norm (by manipulating labor market outcomes or by adjusting housework hours), is present for West Germans but not for East Germans. They explain the differences via the long-lasting effects of the more gender-equal East German institutions. Following the finding that East Germans hold less traditional views regarding the gender division of paid and unpaid work, it is interesting to evaluate whether and how the gender-of-interviewer effect might differ between West and East Germans.<sup>11</sup> The results are presented in Table 7. Separately

<sup>11</sup> East Germany here is defined as the current region of residence at the time of the survey and includes East Berlin but not West Berlin. Due to the geographical definition there is the possibility of within-respondent variation on the *East* dummy. Alternatively, East can be defined based on the biographical information about the residence region before 1989. Results are similar for this definition and available upon request.



**Table 7** Women's housework time—West versus East Germany

	West (1)	East (2)	All (3)	(4)	(5)
Female interviewer	0.077*** (0.03)	0.058 (0.04)	0.078*** (0.03)	0.105*** (0.03)	0.140*** (0.05)
East			0.374 (0.26)	0.252 (0.27)	0.094 (0.29)
Female interviewer * East			−0.011 (0.04)	−0.064 (0.06)	−0.077 (0.10)
Woman full-time				−0.166*** (0.03)	
Woman full-time * East				0.142*** (0.05)	
Female interviewer * woman full-time				−0.070* (0.04)	
Female interviewer * woman full-time * East				0.111* (0.06)	
Relative income woman					−1.235*** (0.13)

Table 7 continued

	West (1)	East (2)	All (3)	(4)	(5)
Relative income woman * East					0.597*** (0.17)
Female interviewer * relative income woman					-0.204* (0.12)
Female interviewer * East * relative income woman					0.211 (0.21)
Adjusted $R^2$	0.107	0.053	0.098	0.099	0.101
Individuals	10462	2748	13145	13145	13145
Observations	47792	13865	61657	61657	61657

Dependent variable is female daily housework hours. The sample is restricted to dual-earner couples (age 18 to 65). The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

studying East German and West German women as in columns 1 and 2, suggests that the gender-of-interviewer effect is smaller and not statistically significant for East German women. The interaction effect between the interviewer's gender and *East* for the whole sample is not statistically significant.

To gain further insight into differences between the groups, we evaluate how the relationship between full-time employment or relative income and the interviewer's gender differs between East and West Germans. Specifically, we include a three-way interaction of the interviewer's gender, the *East* dummy and women's full-time employment in column 4 and women's relative income in column 5. In both specifications, the general gender-of-interviewer effect remains statistically significant and there is no statistically significant interaction effect between *East* and *FemaleInterviewer*. As shown before in Table 6, full-time employed women generally do less housework, for East German women, however, there is a reverse positive interaction effect that is statistically significant. More interesting for our purpose are the interaction effects with the interviewer's gender: While the interaction between the interviewer's gender and women's full-time employment is still negative and now even marginally significant, the three-way interaction with women's full-time employment status is statistically significant and positive. Hence, full-time employed women engage less in gendered reporting, but for East Germans, there is a counteracting effect. The combined effect is not statistically significant (*Coefficient* 0.041, *SE* 0.049, *p-value* 0.403), suggesting that for East German women the gender-of-interviewer effect is not related to the employment status. Regarding the relation to women's relative income, the results in column 5 imply that the negative effect of relative income on housework hours is smaller for East German women. Furthermore, similar to the results in Table 6, the gender-of-interviewer effect is lower when the woman's contribution to the couple's income rises. This negative effect is reversed for East German women, and the combined effect is again not statistically significant (*Coefficient* 0.007, *SE* 0.173, *p-value* 0.967). Hence, for East German women the gender-of-interviewer effect is also not significantly related to women's relative income. Gender-of-interviewer effects and interaction terms with the several characteristics of specialization in paid work remain insignificant for male respondents (see Tables 25 and 26 in the Appendix).

Summarizing, these results support the theory that gender-of-interviewer effects are less present for East German women compared to West German women. Greater attachment to paid work generally seems to reduce gender-of-interviewer effects, but for East German women this relationship is non-existent. This represents an interesting finding as traditional social norms and hence the urge to fulfill a housewife role are less prevalent for East Germans. Additionally, these results further support the conclusion that the established gender-of-interviewer effects are indeed driven by social norms.

## 5.5 Gendered reporting and cultural heritage

Several studies have shown how social norms are related to cultural heritage (Alesina et al. 2013; Giuliano 2020). A recent paper by Blau et al. (2020) studies the role of source country characteristics, especially source country gender equality, in driving the division of non-market work among US immigrants. The major result of the paper

is that the more gender-equal the immigrants' source country, the more equal their gendered division of non-market work. The authors argue that these differences are indeed driven by the internalization of social norms. Hence, source country gender equality of immigrants seems to be a relevant norm-proxy regarding the division of housework. It is therefore interesting to study whether the cultural heritage of immigrants impacts the established gender-of-interviewer effect on housework reports.

To answer this question we follow the approach by Blau et al. (2020) and use the Global Gender Gap Index (GGI) from the World Economic Forum to measure gender norms in source countries. Importantly, the GGI is intentionally designed to be independent of a country's level of development. Instead of focusing on levels of resources or opportunities, the GGI measures gender-based gaps in various indicators of gender equality. Specifically, the index measures the gap between women and men in four categories - economic participation and opportunity, educational attainment, political empowerment, and health and survival - that are combined to create an index with values between zero (total inequality) and 1 (total equality) (Hausmann et al. 2007). Each of these four subindices consists of various indicators of gender equality all measured in female-to-male ratios in access to resources and opportunities.<sup>12</sup>

To capture source country gender equality at a time when immigrants lived in the respective country (and hence were most likely to internalize social norms), it would be ideal to use a GGI from a period prior to immigration. However, GGI has only been measured since 2006. Hence, we follow Blau et al. (2020) and use an average of 2006 and 2007 GGI values for a more stable measure and thereby still predate as many observations as possible.<sup>13</sup> The thus constructed measure of the GGI is matched to all individuals in our sample that do not report Germany as their country of birth. Following the approach by Blau et al. (2020), we further control for source country fertility and GDP per capita to ensure that GGI indeed measures gender equality and is not confounded with a country's level of development.

For the analysis in this section, we focus on the subsample of couples in which both partners are first generation immigrants and originate from the same country.<sup>14</sup> This sample consists of 2,420 immigrant couples from 74 different countries that are observed for an average of 5.5 years, which results in a total of 13,225 observations. The GGI of these observations ranges from the least gender-equal country Pakistan with a GGI of 0.55 to Norway with a GGI of 0.8. The average GGI of all observations is 0.65 compared to the German GGI being 0.75. Hence, female immigrants in our sample on average originate from countries with lower gender equality than Germany.

To study how GGI affects the gender-of-interviewer effect, we include it as an interaction effect with the interviewer's gender in the main specification for the subsample of immigrant couples. For better interpretation, we control for GGI relative to its sam-

<sup>12</sup> The indicators (all measured in female-to-male ratios) cover several measures of wages, labor force participation, literacy rate, school enrollment, healthy life expectancy and sex ratio at birth, ratios of professional and technical employment, as well as seats in parliament, female-to-male ratio at ministerial level, and years as head of state. Each of the several indicators contributing to the subindices is weighted to ensure that it has the same relative impact on the subindex as the other indicators (Hausmann et al. 2007).

<sup>13</sup> For countries that do not have a GGI value in 2006, we use the earliest year available. Results are similar when instead of this approach, the GGI value from 2018 is used for all observations.

<sup>14</sup> Results are similar when immigration is defined only at the individual level.

**Table 8** Housework time and global gender gap index (GGI)

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.120** (0.05)	0.021 (0.02)	0.099* (0.06)
Female interviewer * $[GGI - \overline{GGI}]$	-3.228*** (1.02)	0.010 (0.48)	-3.238*** (1.12)
$[GGI - \overline{GGI}]$	1.110 (0.78)	1.425*** (0.46)	-0.316 (0.92)
Adjusted $R^2$	0.350	0.165	0.351
Observations	13225	13225	13225

Regression for the subsample of immigrant couples in which both partners originate from the same country of origin (age 18 to 65 in cohabiting couples). Dependent variable is daily housework hours of women, men, or within-couple gap. GGI is the Global Gender Gap Index. The table reports the results of OLS regressions. Respondents and Interviewer Controls as well as year and state fixed effects are included. Additional controls are source country fertility and GDP per capita. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

ple mean. Results are presented in Table 8. For these regression analyses, we focus on the cross section, as GGI is constant within couples and otherwise additional source country information like GDP and fertility level could not be included. For the average female immigrant, the gender-of-interviewer effect persists, suggesting that female immigrants report more housework hours toward female interviewers. Interestingly, the interaction term with GGI is negative: The gender-of-interviewer effect is *lower* for individuals from *more* gender-equal countries. Or explained differently, female immigrants from more gender-equal countries engage less in gendered reporting, while female immigrants from less gender-equal countries do so more. This result further supports the hypothesis that a stronger internalization of traditional social norms leads to a larger gender-of-interviewer effect. The average male immigrant does not engage in gendered reporting and the gender-of-interviewer effect does not depend on the GGI. However, male immigrants from more gender-equal countries appear to do more housework. As there are no significant gender-of-interviewer effects for males, the within-couple housework gap increases when the interviewer is female and decreases for immigrants from more gender-equal countries.

In Table 27 in the Appendix, we demonstrate that results are mostly unaffected when additionally controlling for the duration of immigrants' residence in Germany. Stratifying the sample into two groups – women who have lived in Germany for less than or at least 20 years – shows that the relation between the gender-of-interviewer effect and GGI is statistically significant and of similar magnitude in both groups. Thus, independent of the duration of immigrants' stay in Germany, the gender-of-interviewer effect is significantly smaller for immigrants from more gender-equal countries. The gender-of-interviewer effect for the *average* female immigrant turns insignificant for those who have resided in Germany for 20 years or longer, which could be some

indication for cultural assimilation. For male respondents the gender-of-interviewer effect and its interaction with GGI remain insignificant, as in Table 8.

## 6 Gender versus alternative explanations

The results presented in this paper are evidence of a causal effect of the interviewer's gender on women's reporting of housework hours which we interpret as an indication for the role of social norms. To address potential concerns about other factors than the interviewer's gender driving the results, this section provides several robustness checks.

As explained above, female and male interviewers differ in characteristics that may affect interview responses besides the interviewer's gender. We still do not include these as control variables in the main specification because most variables are only collected in three specific years. Including these variables results in dropping all interviewers that were not present in at least one of those years, which reduces the variation in the interviewer's gender and creates gaps in the data. To alleviate concerns that our results are driven by other interviewer characteristics than their gender, we rerun the main regressions as in equation (1) additionally controlling for several of these interviewer characteristics. More specifically, we include dummy variables reflecting whether the interviewer is married, has an upper secondary degree, is born in Germany and whether the interviewer speaks German as his or her mother tongue. Results are presented in Table 9.

The results show that almost none of the additional interviewer characteristics seem to affect the housework responses. The only statistically significant effect is associated with men reporting fewer housework hours when they are confronted with an interviewer of a higher educational degree. Apart from that, the coefficients for the interviewer's gender are even larger than in the main specification. This, however, does not seem to reflect the importance of the additional interviewer characteristics, but rather the much smaller and more selective sample that results from their inclusion.<sup>15</sup> That the effect of the interviewer's gender not only remains stable but is larger in these alternative specifications supports the stability of the established effect.

As argued in section 4, the only concern regarding the causality of the established gender-of-interviewer effect arises if a changed interviewer's gender goes along with a parallel change of unobserved factors that also affect housework hours. In line with this, one potential concern with our analysis regards the timing of female versus male interviewers both across years and within a year, as both are potentially also related to the performing of housework hours: For instance, assume that over the observed time period the share of female interviewers decreases substantially, so that most interviewer-gender-transitions over time are from female-to-male interviewers and not in the opposite direction. As over the years, female housework hours on average also decreased, the established effect of larger housework reports toward

<sup>15</sup> Performing the regression analyses for identical samples but once with and once without the additional interviewer characteristics yields fairly similar results. This further supports the conclusion that the interviewer-gender-effect differs because of the selectiveness of the samples and not because of the relevance of other interviewer characteristics.

**Table 9** Housework time and other interviewer characteristics

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.163*** (0.06)	0.049 (0.03)	0.114* (0.07)
Age	−0.001 (0.00)	−0.002 (0.00)	0.001 (0.00)
Married	−0.040 (0.03)	0.001 (0.02)	−0.041 (0.04)
Higher education	−0.044 (0.03)	−0.048** (0.02)	0.004 (0.04)
Born in Germany	−0.039 (0.15)	−0.138 (0.11)	0.099 (0.19)
Mother tongue German	−0.679 (0.53)	0.036 (0.17)	−0.714 (0.54)
Adjusted $R^2$	0.112	0.089	0.131
Individuals	8803	8803	8803
Observations	45309	45309	45309

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHoursWoman* − *HouseworkHoursMan*). The table reports the results of individual (or couple) fixed effects estimations. Additional controls include paid work time, income, age, age squared, education categories for women and men respectively, marital status, household income and number of children in three different age groups, interview mode, as well as year and state fixed effects. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

female interviewers might then simply be driven by the coexistence of these two developments. This concern can, however, be alleviated by analyzing the distribution of interviews with female interviewers over time (see Fig. 2 in the Appendix). While the share of interviews being conducted by female interviewers varies between 37% and 47%, it does not exhibit a clear trend over time. Hence, it seems very unlikely that the aforementioned mechanism is impacting our results.

The second source of time variation that potentially affects our results stems from variation within a given year: If female interviewers are conducting more interviews at times in the year when women also do more housework or vice versa, it could theoretically be that these two independent factors lead to our results. For instance, it may be that female interviewers are conducting fewer interviews in the summer months (e.g., because of their childrens' school breaks) and women also perform less housework during that time.<sup>16</sup> To analyze potential variation over the year in Fig. 3 in the Appendix we illustrate the share of women's interviews with female interviewers over months of the year. While there is indeed some variation over the months, it does not seem to be large enough to allow for the reasoning described above.

<sup>16</sup> Note that the questionnaire asks for housework on a "normal workday" which makes it less likely that estimates are influenced by seasonality, at least compared to questions asking for time-use in the last week.



Additionally, the gender-of-interviewer effect is stable when including the month of interview dummies as control variables (see Table 28 in the Appendix).

Besides other interviewer characteristics and timing of the interviews, a potential concern is that the established effect may not be driven by the interviewer's gender, but instead by the change in the interviewer per se. It is possible, for example, that respondents and interviewers establish a relationship of trust over the years of repeated interviews which leads to more truthful reporting (Kühne 2018). Then, a change in the interviewer would affect the reporting irrespective of the interviewer's gender. To address this concern we build a placebo test, where we analyze changes in the interviewer irrespective of interviewer-gender changes. Hence, we compare interviewer changes with situations where the interviewer does not change.<sup>17</sup> Results of this regression are presented in the first and second column of Table 10. Column 1 shows that a general change in the interviewer, irrespective of a gender change, does significantly increase the housework reported by female respondents while there is no effect for males. However, both size and statistical significance of the gender-of-interviewer effect are almost unaffected by additionally controlling for a change in the interviewer. This is confirmed by the results in column 2 in which an interaction effect between the interviewer's gender and interviewer change is included. In all years in which there is no change in the interviewer, women report more housework to female as opposed to male interviewers. The effect of a transition from male to female interviewer is given by the combined effect of all three terms (Coefficient 0.121, SE.031, p-value 0.000). Thus, there is an increase in housework reports in the first year in which respondents experience the transition from male to female interviewer. Which part of this effect is driven by the change per se, and which by the interviewer gender change cannot be disentangled. Overall these results imply that there is some effect when respondents experience a change in the interviewer, however, the established gender-of-interviewer effect persists regardless.

An additional test to check the robustness of our results derives from analyzing different interview modes. While most interviews are indeed face-to-face interviews, in around 24% of the interviews in our sample the respondent fills in the questionnaire by him- or herself with pencil and paper. Even in these cases, however, an interviewer is present in the household. If our results are driven by the interaction between respondent and interviewer, we would expect that effects are smaller or absent when respondents fill in the questionnaire on their own. As shown in Table 10 it is indeed the case that the effect is smaller and not statistically significant for female respondents who do not directly interact with an interviewer. As the interviewer is present in the household in both cases, the difference between individuals who self fill in versus those who do not, again supports the idea that interviewer gender is not correlated to other unobserved factors but indeed drives the results. For male respondents, there is no gender-of-interviewer effect for those who do not fill in the questionnaire by themselves, but there appears to be some weakly significant effect for those without a direct interaction with the interviewer. This effect, however, is likely driven by the selectivity of the restricted sample which contains gaps in the data, as the effect also disappears in a cross-sectional

<sup>17</sup> Specifically, we create a dummy variable coded one for each year individuals face a new interviewer and zero in all other years. The first observed interview year for all individuals is also coded zero, so that the dummy variable only captures *changes* in the interviewer.

**Table 10** Housework time and Placebo tests

	Placebo change		No Self fill in (3)	Interview mode Self fill in (4)	No mode change (5)
	(1)	(2)			
<i>Women</i>					
Female interviewer	0.079*** (0.02)	0.089*** (0.02)	0.058** (0.03)	0.031 (0.04)	0.085*** (0.03)
Interviewer change	0.063*** (0.02)	0.086*** (0.02)			
Female interviewer * interviewer change		−0.055 (0.04)			
Adjusted $R^2$	0.167	0.167	0.146	0.195	0.135
<i>Men</i>					
Female interviewer	0.014 (0.01)	0.013 (0.01)	0.025 (0.02)	−0.043* (0.02)	−0.000 (0.02)
Interviewer change	0.015 (0.01)	0.013 (0.01)			
Female interviewer * Interviewer change		0.004 (0.02)			
Adjusted $R^2$	0.097	0.097	0.094	0.098	0.087
Individuals	18965	18965	15572	5192	14497
Observations	110544	110544	70928	26418	58866

Dependent variable is female or male daily housework hours. The table reports the results of individual fixed effects estimations. Respondents and Interviewer Controls as well as year and state fixed effects are included. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

analysis (not shown here). Importantly, the gender-of-interviewer effect remains stable when only including couples who never have changes in the interview mode, which further alleviates concerns about the interview mode affecting the results.

## 7 Conclusion

A growing body of literature seeks to understand how social norms and social roles affect behavior. In this field of research, the gendered division of housework hours has received increasing attention, not only by researchers but also in the public debate. Our results present evidence for the fact that social norms impact response behavior in interviews. We find a causal effect of the interviewer's gender on women's housework reporting: Women in couple relationships report more time spent on housework when they are interviewed by a woman instead of a man. We do not find any gender-of-interviewer effects for male respondents, which leads to an observed interviewer gender gap in the housework gender gap. Therefore, estimates about time-use from face-to-face interviews should be interpreted with caution. We argue that the observed gender-of-interviewer effect is most likely driven by self-deceptive reporting behavior. Women internalize societal expectations more strongly when they interact with a female as opposed to a male interviewer, which subsequently affects their housework report. Stratifying the sample and interacting the interviewer's gender with several norm-related characteristics support the hypothesis that the observed gender-of-interviewer effects are driven by social norms.

Note, while our results provide strong evidence for the existence of gendered reporting in time-use estimates, our findings do not imply that interviewer effects are the only way in which social norms affect the reporting in surveys. By comparing responses toward female and male interviewers, our approach captures only a part of potential adjustments in the reporting process. It is possible that social norms affect the reports beyond what is captured by the reaction to the interviewer's gender. Importantly, it remains an open question whether even diary-based estimates of time-use are influenced by the social roles individuals try to fulfill. As there are no data on the "true" time-use values, it is impossible to compare interview reports to reality. Increasing the use of mixed-method approaches in survey design could be beneficial for testing differences between various modes concerning outcomes on a wide range of topics. While face-to-face interviews are known to have significant advantages compared to other interview modes (such as building trust on sensitive topics), our study implies that drawbacks may exist even for topics not previously considered as potentially affected by interviewer effects.


More broadly, our findings offer a new perspective on the discussion about norms influencing behavior. While previous studies on socially desirable reporting have almost exclusively focused on normative views (e.g., about gender equality) or behaviors clearly linked to societal expectations (such as drug use or voting behavior), our study provides evidence that socially desirable reporting matters even for topics less clearly linked to social norms. Therefore, our understanding of the impact of norms on behavior may be limited by the prevalence of reporting biases for all kinds of outcomes. As a consequence, future research should take into account the conceivably

strong effects of reporting behavior in domains not previously regarded as relevant to this issue.

Appendix

See Figs. 1, 2, 3 and Tables 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27 and 28 in the appendix

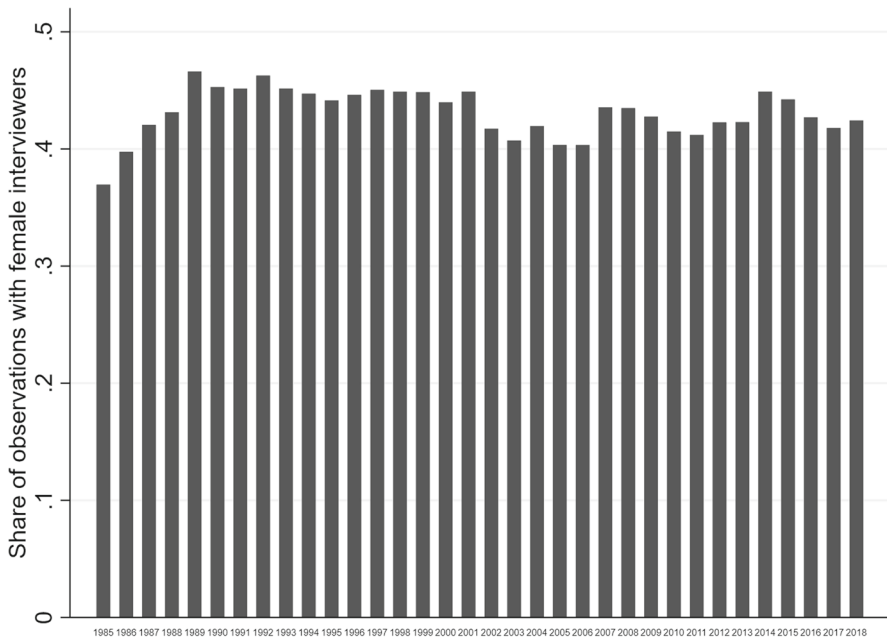
**What is a typical weekday like for you? How many hours per normal workday do you spend on the following activities?**

 Please give only whole hours.  
Use zero if the activity does not apply!

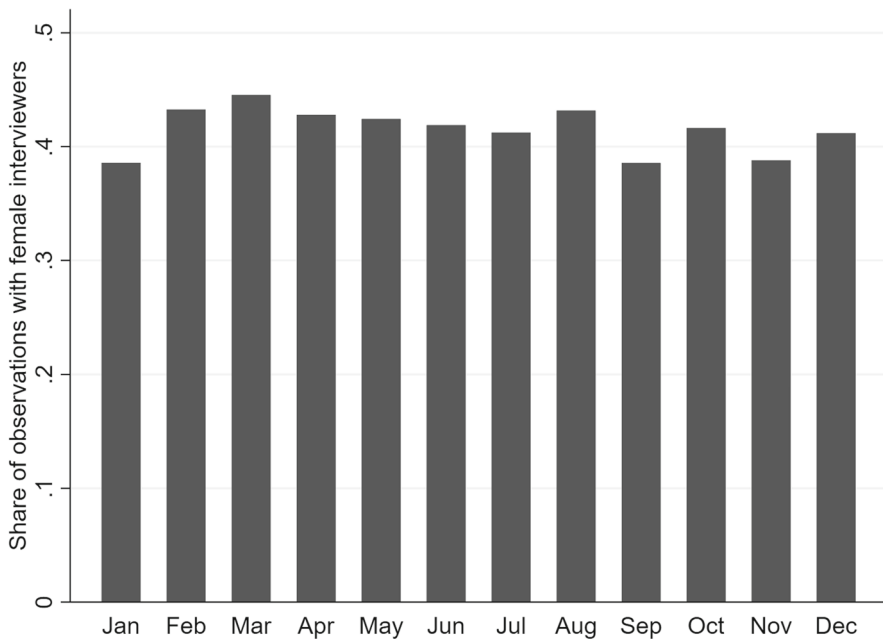
Number of hours

Job, apprenticeship, second job (including travel time to and from work) .....	<div></div>	<div></div>
Errands (shopping, trips to government agencies, etc.).....	<div></div>	<div></div>
Housework (washing, cooking, cleaning) .....	<div></div>	<div></div>
Child care .....	<div></div>	<div></div>
Care and support for persons in need of care.....	<div></div>	<div></div>
Education or further training (also school, university) .....	<div></div>	<div></div>
Repairs on and around the house, car repairs, garden work .....	<div></div>	<div></div>
Physical activities (sports, fitness, gymnastics).....	<div></div>	<div></div>
Other leisure activities and hobbies .....	<div></div>	<div></div>

**Fig. 1** Interview question - time-use. *Time-use question of the personal interview questionnaire - SOEP 2018*



**Fig. 2** Share of female interviewers over time. *The Figure is based on the main sample as defined in section 3. The bars represent the share of interviews with female interviewers over time*



**Fig. 3** Share of female interviewers by interview month. *The Figure is based on the main sample as defined in section 3. The bars represent the share of interviews with female interviewers by interview month. In less than 4 percent of cases interviews of partners were in a different month. For simplicity here only interviews with female respondents are shown*

**Table 11** Housework time and interviewer gender - same day interview

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.080*** (0.02)	0.012 (0.01)	0.068*** (0.03)
Adjusted $R^2$	0.165	0.100	0.169
Individuals	18149	18149	18149
Observations	97176	97176	97176

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65) but without couples who were interviewed at different dates. Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHours Woman – HouseworkHours Man*). The table reports the results of individual fixed effects estimations. Additional controls as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 12** Housework time and interviewer gender - restricting housework hours

	Housework hours			
	$\leq 16$ (1)	$\leq 12$ (2)	$\leq 8$ (3)	$\leq 6$ (4)
<i>Dep. Var: Female housework hours</i>				
Female interviewer	0.079*** (0.02)	0.077*** (0.02)	0.063*** (0.02)	0.056*** (0.02)
Adjusted $R^2$	0.170	0.169	0.170	0.161
Individuals	18933	18857	17937	16465
Observations	110142	109246	100555	87413
<i>Dep. Var: Male housework hours</i>				
Female interviewer	0.014 (0.01)	0.016 (0.01)	0.018 (0.01)	0.013 (0.01)
Adjusted $R^2$	0.097	0.097	0.102	0.099
Individuals	18965	18961	18904	18825
Observations	110544	110509	109926	109170

The sample contains only cohabiting couples (age 18–65). Control variables as in the main specification. In all regressions we only keep observations who never exceed the respective threshold of housework hours. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 13** Housework time and interviewer gender - interviewer experience

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.078*** (0.02)	0.015 (0.01)	0.063*** (0.02)
Interviewer experience	0.002 (0.00)	−0.001 (0.00)	0.003 (0.00)
Adjusted $R^2$	0.167	0.097	0.169
Individuals	18965	18965	18965
Observations	110544	110544	110544

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHours Woman* – *HouseworkHours Man*). The table reports the results of individual (or couple) fixed effects estimations. Interviewer Experience measures the cumulative number of years interviewers have worked for SOEP. Additional controls as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$



**Table 14** Respondent characteristics by interviewer gender

	Mean male interviewer	Mean female interviewer	Diff.	SE diff.
Income man	2029.94	2087.38	−57.434***	10.484
Income woman	781.22	767.31	13.904*	5.518
Income HH	3437.45	3461.91	−24.467	12.534
Paid work time man	36.48	36.98	−0.493***	0.114
Paid work time woman	19.54	19.38	0.157	0.110
Man's age	45.45	45.18	0.273***	0.065
Woman's age	42.70	42.41	0.293***	0.065
Education level man	5.01	5.02	−0.011	0.015
Education level woman	4.94	4.89	0.053***	0.014
Married	0.88	0.88	0.002	0.002
No. of children age <6	0.30	0.30	−0.006	0.004
No. of children age 6-12	0.43	0.43	−0.002	0.004
No. of children age 13-16	0.23	0.23	0.003	0.003
Interv. day man	95.22	94.75	0.470	0.350
Interv. day woman	94.96	94.27	0.693*	0.350
Interview length man	35.10	36.40	−1.300***	0.081
Interview length woman	34.21	35.39	−1.185***	0.080
Observations	110,544			

The table reports mean comparisons of respondents' characteristics by interviewer gender based on the main sample (cohabiting couples of age 18 to 65). Paid work time is hours worked per week including overtime hours. Income is net monthly income and household income is net monthly income of all household members including non-labor income. Income is corrected for inflation with base year 2016. Education levels derive from the CASMIN classification. Married is a dummy variable equal to 1 for married couples. No. of children is the number of children living in the household for each of the age groups. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 15** Housework time and interviewer gender - cross section

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.078*** (0.02)	0.009 (0.01)	0.069*** (0.02)
<i>Respondent Characteristics</i>			
Paid work time woman	-0.037*** (0.00)	0.007*** (0.00)	-0.044*** (0.00)
Paid work time man	0.008*** (0.00)	-0.018*** (0.00)	0.026*** (0.00)
Income woman	-0.000*** (0.00)	0.000*** (0.00)	-0.000*** (0.00)
Income man	-0.000* (0.00)	-0.000*** (0.00)	0.000** (0.00)
Log. HH income	-0.010 (0.03)	-0.022 (0.02)	0.012 (0.03)
Married	0.169*** (0.02)	-0.069*** (0.01)	0.238*** (0.02)
No. of children age <6	0.352*** (0.02)	0.090*** (0.01)	0.262*** (0.02)
No. of children age 6-12	0.325*** (0.01)	0.040*** (0.01)	0.286*** (0.01)
No. of children age 13-16	0.325*** (0.01)	0.026*** (0.01)	0.299*** (0.02)
<i>Interviewer characteristics</i>			
Age	0.001 (0.00)	0.000 (0.00)	0.001 (0.00)
Interview length woman	0.002*** (0.00)	0.001** (0.00)	0.001 (0.00)
Interview length man	0.003*** (0.00)	0.001*** (0.00)	0.002*** (0.00)
Adjusted $R^2$	0.415	0.189	0.406
Observations	110544	110544	110544

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHours Woman - HouseworkHours Man*). The table reports the results in the cross section without controlling for individual fixed effects. Additional controls include age, age squared, education categories for women and men, respectively, interview mode, as well as year and state fixed effects as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 16** Housework time and interviewer gender - exogenous controls

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.082*** (0.02)	0.010 (0.01)	0.072*** (0.03)
<i>Respondent characteristics</i>			
Married	0.223*** (0.03)	−0.094*** (0.02)	0.317*** (0.04)
No. of children age <6	0.694*** (0.02)	0.000 (0.01)	0.693*** (0.02)
No. of children age 6–12	0.384*** (0.02)	0.020** (0.01)	0.364*** (0.02)
No. of children age 13–16	0.227*** (0.02)	0.023*** (0.01)	0.203*** (0.02)
<i>Interviewer characteristics</i>			
Age	−0.001 (0.00)	0.001 (0.00)	−0.002 (0.00)
Interview length woman	−0.003*** (0.00)	0.002*** (0.00)	−0.004*** (0.00)
Interview length man	0.004*** (0.00)	−0.001*** (0.00)	0.005*** (0.00)
Adjusted $R^2$	0.095	0.024	0.070
Individuals	18965	18965	18965
Observations	110544	110544	110544

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHours Woman* − *HouseworkHours Man*). The table reports the results of individual fixed effects estimations. Additional controls include age, age squared, education categories for women and men, respectively, interview mode, as well as year and state fixed effects as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 17** Total non-market work and interviewer gender

	Household chores		Total non-market work	
	Women (1)	Men (2)	Women (3)	Men (4)
Female interviewer	0.058* (0.03)	0.024 (0.02)	0.208*** (0.08)	0.040 (0.05)
Adjusted $R^2$	0.130	0.146	0.306	0.135
Individuals	17549	17548	14459	14462
Observations	98965	98875	72060	72017

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily hours for women and men for the sum of household chores (housework, repairs/gardening, errands) in columns 1 and 2 and for the sum of all non-market activities (chores, childcare, elderly care) in columns 3 and 4. The table reports the results of individual fixed effects estimations. Additional controls include age, age squared, education categories for women and men respectively, interview mode, as well as year and state fixed effects. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 18** Gender-of-interviewer effects for other activities

	Childcare (1)	Repair/garden (2)	Elderly care (3)	Errands (4)	Education (5)	Sleep (6)	Work hours (7)	Exercise (8)	Leisure (9)
<i>Women</i>									
Female interviewer	0.142*** (0.05)	0.022* (0.01)	0.002 (0.02)	0.006 (0.01)	-0.015 (0.01)	-0.006 (0.03)	0.031 (0.05)	0.015 (0.02)	0.053** (0.03)
Adjusted $R^2$	0.332	0.029	0.004	0.035	0.102	0.017	0.198	0.009	0.066
Individuals	18963	18959	14468	17554	18949	9369	18909	7675	18944
Observations	110521	110503	72124	98999	110399	31798	109138	23973	108560
<i>Men</i>									
Female interviewer	0.008 (0.02)	0.009 (0.01)	0.005 (0.01)	0.010 (0.01)	0.006 (0.01)	0.001 (0.02)	0.044 (0.04)	0.005 (0.02)	0.058** (0.03)
Adjusted $R^2$	0.101	0.057	0.003	0.059	0.099	0.017	0.402	0.014	0.108
Individuals	18965	18961	14469	17551	18957	9364	18900	7674	18938
Observations	110531	110413	72128	98993	110374	31797	109304	23971	108524

Dependent variable is daily hours for different activities. The sample is restricted to couples aged 18 to 65. The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included as in the main specification. For the regressions with work hours as dependent variable we exclude actual work hours as explanatory variable (it is not identical to the variable used in the main specification as here e.g. travel time is included but highly collinear). Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 19** Logarithm of housework hours - relative effects (couples and singles)

	Women couples (1)	Men couples (2)	Single women (3)	Single men (4)
Female interviewer	0.024*** (0.01)	0.006 (0.01)	-0.012 (0.01)	0.001 (0.01)
Adjusted $R^2$	0.192	0.079	0.108	0.048
Individuals	18873	15260	7487	4606
Observations	108794	61231	32793	18326

Dependent variable is the logarithm of female or male daily housework hours. The sample is restricted to individuals or couples aged 18 to 65. The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 20** Men's housework time - respondent and interviewer cohorts

	Young cohort (1)	Old cohort (2)	All (3)	Young cohort (4)	Old cohort (5)
Female interviewer	0.011 (0.01)	0.017 (0.02)	0.016 (0.02)	0.013 (0.02)	0.019 (0.03)
[Reference: interviewer born < 1945] Interviewer born 1945-1959			-0.018 (0.03)	-0.027 (0.03)	-0.015 (0.04)
Interviewer born >= 1960			-0.023 (0.05)	-0.066 (0.06)	0.041 (0.08)
Female interviewer			-0.004 (0.03)	-0.012 (0.03)	0.009 (0.05)
* Interviewer born 1945-1959			0.015 (0.04)	0.044 (0.05)	-0.059 (0.08)
Female interviewer					
* Interviewer born >= 1960					
Adjusted $R^2$	0.084	0.110	0.097	0.084	0.109
Individuals	12027	6938	18965	12027	6938
Observations	63757	46787	110544	63757	46787

Dependent variable is male daily housework hours. The sample is restricted to cohabiting couples (age 18 to 65). Columns are based on respondent cohorts: Young cohort is defined as at least one person in the couple being born 1960 or later. The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 21** Gender-of-interviewer effects and interviewer's age

	Woman's Housework Time (1)	Man's Housework Time (2)
Female interviewer	0.082*** (0.02)	0.014 (0.01)
<i>Interviewer Age</i> – $\overline{\text{Interviewer Age}}$	–0.001 (0.00)	0.000 (0.00)
Female interviewer * [ <i>Interviewer Age</i> – $\overline{\text{Interviewer Age}}$ ]	0.003 (0.00)	0.000 (0.00)
Adjusted $R^2$	0.167	0.097
Individuals	18965	18965
Observations	110544	110544

Dependent variable is female or male daily housework hours. The sample is restricted to cohabiting couples (age 18 to 65). Interviewer age is demeaned for better interpretation. The table reports the results of individual fixed effects estimations. Respondents and Interviewer controls as well as year and state fixed effects are included as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$



**Table 22** Housework time and interviewer gender - by respondent education

	Women		Men		All (6)
	High education (1)	Low education (2)	High education (4)	Low education (5)	
Female interviewer	0.088** (0.04)	0.078*** (0.03)	0.014 (0.02)	0.018 (0.01)	0.013 (0.01)
High educ.					0.140*** (0.04)
Female interviewer * high educ.					0.007 (0.02)
Adjusted $R^2$	0.176	0.168	0.100	0.098	0.097
Individuals	3684	15486	4588	14593	18965
Observations	20791	89753	25852	84692	110544

Dependent variable is female or male daily housework hours. The sample is restricted to cohabiting couples aged 18 to 65. High education is defined as lower or higher tertiary education with the CASMIN-classification. The table reports the results of individual fixed effects estimations. Respondents and interviewer controls as well as year and state fixed effects are included as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 23** Housework time and interviewer gender - by presence of children

	Women (1)	Men (2)
Female interviewer	0.082*** (0.03)	0.021 (0.01)
[Reference: no children in HH]		
Children age < 6	0.435*** (0.03)	0.052*** (0.02)
Children age ≥ 6 & ≤ 16	0.222*** (0.02)	0.017 (0.01)
Female interviewer * children age < 6	0.008 (0.04)	−0.011 (0.02)
Female interviewer * children age ≥ 6 & ≤ 16	−0.022 (0.03)	−0.015 (0.02)
Adjusted $R^2$	0.161	0.097
Individuals	18965	18965
Observations	110544	110544

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65). Dependent variable is daily housework hours of women or men. The table reports the results of individual fixed effects estimations. Additional controls as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 24** Housework time and interviewer gender by employment status - extensive margin

	Women not employed (1)	Women employed (2)	All women (3)
Female interviewer	0.103** (0.05)	0.080*** (0.02)	0.071** (0.03)
Woman employed			−0.321*** (0.03)
Female interviewer * woman employed			0.013 (0.03)
Adjusted $R^2$	0.088	0.106	0.169
Individuals	10884	14253	18965
Observations	40170	70374	110544

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65) differentiating by employment status for women. Dependent variable is daily housework hours of women. The table reports the results of individual fixed effects estimations. Additional controls as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 25** Housework time interactions - male housework hours

	Full-time (1)	Employment status Not full-time (2)	All (3)	Male higher or equal (4)	Relative income Female Higher (5)	All (6)	All (7)
Female interviewer	0.003 (0.02)	-0.010 (0.02)	-0.005 (0.02)	0.014 (0.01)	0.006 (0.04)	0.004 (0.01)	-0.009 (0.02)
Female interviewer * Woman full-time			0.017 (0.02)				
Woman full-time			0.010 (0.02)				
Female interviewer * Woman earns more						-0.004 (0.02)	
Female interviewer * Relative income woman							0.035 (0.06)
Woman earns more						0.002 (0.02)	0.001 (0.02)
Relative income woman						0.567*** (0.08)	0.552*** (0.08)
Adjusted $R^2$	0.027	0.028	0.031	0.030	0.031	0.033	0.033
Individuals	7032	8890	13145	11861	3349	13145	13145
Observations	26695	34962	61657	51779	9878	61657	61657

Dependent variable is male daily housework hours. The sample is restricted to dual-earner couples (age 18 to 65). The table reports the results of individual fixed effects estimations. Respondents and interviewer controls as well as year and state fixed effects are included. Full-time and relative income are still defined for the female person in the couple for better interpretation. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 26** East interactions - male housework hours

	West (1)	East (2)	All (3)	(4)	(5)
Female interviewer					
East	-0.000 (0.02)	-0.003 (0.02)	0.001 (0.02)	-0.009 (0.02)	-0.024 (0.03)
			0.275** (0.12)	0.281** (0.12)	0.338*** (0.13)
Female interviewer * East			0.004 (0.03)	0.027 (0.03)	0.091 (0.06)
Female interviewer * woman full-time				0.028 (0.02)	
Woman full-time				0.012 (0.02)	
Woman full-time * East				-0.003 (0.03)	
Female interviewer * woman full-time * East				-0.046 (0.04)	
Female interviewer * relative income woman					0.083 (0.08)
Relative income woman					0.589*** (0.09)

Table 26 continued

	West (1)	East (2)	All (3)	(4)	(5)
Relative income woman * East					-0.135 (0.11)
Female interviewer * East * relative income woman					-0.233 (0.15)
Adjusted $R^2$	0.037	0.012	0.031	0.031	0.033
Individuals	10462	2748	13145	13145	13145
Observations	47792	13865	61657	61657	61657

Dependent variable is male daily housework hours. The sample is restricted to dual-earner couples (age 18 to 65). The table reports the results of individual fixed effects estimations. Respondents and interviewer controls as well as year and state fixed effects are included. Full-time and relative income are still defined for the female person in the couple for better interpretation. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 27** Housework time and global gender gap index (GGI) - years spent in Germany

	Women		Men	
	Years in Germany <20 (1)	Years in Germany >=20 (2)	Years in Germany <20 (3)	Years in Germany >=20 (4)
Female interviewer	0.122** (0.05)	0.169*** (0.06)	0.048 (0.08)	0.030 (0.02)
Female interviewer * $[GGI - \overline{GGI}]$	-3.136*** (1.02)	-3.200*** (1.21)	-3.409** (1.58)	-0.133 (0.46)
$[GGI - \overline{GGI}]$	1.201 (0.82)	1.570 (1.02)	0.814 (1.28)	1.352*** (0.47)
Years in Germany	0.005 (0.00)			-0.004** (0.00)
Adjusted $R^2$	0.350	0.374	0.326	0.165
Observations	12813	7298	5515	12826

Regression for the subsample of immigrant couples in which both partners originate from the same country of origin (age 18 to 65 in cohabiting couples). Dependent variable is daily housework hours of women, men, or within-couple gap. Couples 1-3 refer to female respondents, and columns 4-6 to male respondents. GGI is the Global Gender Gap Index. The table reports the results of OLS regressions. Respondents and Interviewer Controls as well as year and state fixed effects are included. Additional controls are source country fertility and GDP per capita. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Table 28** Housework time and interviewer gender - control for interview month

	Women (1)	Men (2)	Gap (3)
Female interviewer	0.080*** (0.02)	0.011 (0.01)	0.069*** (0.03)
February	0.008 (0.02)	−0.024** (0.01)	0.032 (0.02)
March	0.037 (0.02)	−0.040*** (0.01)	0.077*** (0.03)
April	0.043* (0.02)	−0.047*** (0.01)	0.090*** (0.03)
May	0.034 (0.03)	−0.030* (0.02)	0.064** (0.03)
June	0.063** (0.03)	−0.027 (0.02)	0.091*** (0.03)
July	0.082** (0.03)	−0.049*** (0.02)	0.131*** (0.04)
August	0.069* (0.04)	−0.030 (0.02)	0.098** (0.04)
September	0.143*** (0.04)	−0.020 (0.02)	0.162*** (0.05)
October	0.143** (0.06)	−0.059* (0.03)	0.202*** (0.07)
November	0.081 (0.12)	0.041 (0.06)	0.041 (0.14)
December	0.207 (0.35)	0.157 (0.11)	0.050 (0.38)
Adjusted $R^2$	0.166	0.100	0.170
Individuals	18149	18149	18149
Observations	97176	97176	97176

The table reports regression analyses for the main sample (cohabiting couples of age 18 to 65) but without couples who were interviewed at different dates. Dependent variable is daily housework hours of women, men or within-couple gap (*HouseworkHours Woman* − *HouseworkHours Man*). The table reports the results of individual fixed effects estimations. Additional controls as in the main specification. Standard errors clustered at the individual level are in parentheses. Significance is denoted by \*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Data Availability** The SOEP data are not publicly available but researchers can sign a data distribution contract to access the data, <https://doi.org/10.5684/soep-core.v35>.

The index of gender equality (GGI) comes from the World Economic Forum's "The Global Gender Gap Report", available at [https://www3.weforum.org/docs/WEF\\_GenderGap\\_Report\\_2007.pdf](https://www3.weforum.org/docs/WEF_GenderGap_Report_2007.pdf).

Total fertility and GDP data come from the World Bank, available at <https://data.worldbank.org/indicator/SP.DYN.TFRT.IN> and <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>.

## Declarations

**Competing interests** The authors have no financial or non-financial interests to disclose that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akerlof GA, Kranton RE (2000) Economics and identity. *Q J Econ* 115(3):715–753
- Alesina A, Fuchs-Schündeln N (2007) Good-bye Lenin (or not?): the effect of communism on people's preferences. *Am Econ Rev* 97(4):1507–1528
- Alesina A, Giuliano P, Nunn N (2013) On the origins of gender roles: women and the plough. *Q J Econ* 128(2):469–530
- Bauernschuster S, Rainer H (2012) Political regimes and the family: how sex-role attitudes continue to differ in reunified Germany. *J Popul Econ* 25(1):5–27
- Beblo M, Görges L (2018) On the nature of nurture. The malleability of gender differences in work preferences. *J Econ Behav Organ* 151:19–41
- Beblo M, Robledo JR (2008) The wage gap and the leisure gap for double-earner couples. *J Popul Econ* 21(2):281–304
- Bertrand M, Kamenica E, Pan J (2015) Gender identity and relative income within households. *Q J Econ* 130(2):571–614
- Bittman M, England P, Sayer L, Folbre N, Matheson G (2003) When does gender trump money? Bargaining and time in household work. *Am J Sociol* 109(1):186–214
- Blau FD, Kahn LM, Comey M, Eng A, Meyerhofer P, Willén A (2020) Culture and gender allocation of tasks: source country characteristics and the division of non-market work among US immigrants. *Rev Econ Househ* 18(4):907–958
- Bohlender A, Rathje M, Glemser A (2020) SOEP-Core - 2018. Report of Survey Methodology and Field-work
- Chapman CD, Benedict C, Schiöth HB (2018) Experimenter gender and replicability in science. *Sci Adv* 4(1):e1701427
- Coltrane S (2000) Research on household labor: modeling and measuring the social embeddedness of routine family work. *J Marriage Fam* 62(4):1208–1233
- Cooke LP (2007) Persistent policy effects on the division of domestic tasks in reunified Germany. *J Marriage Fam* 69(4):930–950
- Deutsches Institut für Wirtschaftsforschung (2020). Socio-Economic Panel (SOEP), data for years 1984–2018: version 35, SOEP
- Fernández R (2013) Cultural change as learning: the evolution of female labor force participation over a century. *Am Econ Rev* 103(1):472–500
- Flores-Macias F, Lawson C (2008) Effects of interviewer gender on survey responses: findings from a household survey in Mexico. *Int J Pub Opin Res* 20(1):100–110
- Foster G, Stratton LS (2018) Do significant labor market events change who does the chores? Paid work, housework, and power in mixed-gender Australian households. *J Popul Econ* 31(2):483–519
- Geist C (2009) One Germany, two worlds of housework? Examining employed single and partnered women in the decade after unification. *J Comp Fam Stud* 40(3):415–437
- Gershuny J (2013) National utility: measuring the enjoyment of activities. *Eur Sociol Rev* 29(5):996–1009
- Giuliano P (2020) Gender and culture. *Oxf Rev Econ Policy* 36(4):944–961



- Goebel J, Grabka MM, Liebig S, Kroh M, Richter D, Schröder C, Schupp J (2019) The German socio-economic panel (SOEP). *Jahrb Natl Stat* 239(2):345–360
- Hausmann R, Tyson L D, Zahidi S (2007) The Global Gender Gap Report 2007
- Holtgraves T (2004) Social desirability and self-reports: testing models of socially desirable responding. *Pers Soc Psychol Bull* 30(2):161–172
- Huddy L, Billig J, Bracciodieta J, Hoeffler L, Moynihan PJ, Pugliani P (1997) The effect of interviewer gender on the survey response. *Polit Behav* 19(3):197–220
- Jessen J, Schweighofer-Kodritsch S, Weinhardt F, Berkes J (2024) Separate housework spheres. *IZA DP No.* 17134
- Kan MY (2008) Measuring housework participation: the gap between “stylised” questionnaire estimates and diary-based estimates. *Soc Indic Res* 86(3):381–400
- Kane EW, Macaulay LJ (1993) Interviewer gender and gender attitudes. *Pub Opin Q* 57(1):1–28
- Kantar Public (2019). SOEP-Core - 2018. Personenfragebogen, Stichproben A-L3+N
- Krumpal I (2013) Determinants of social desirability bias in sensitive surveys: a literature review. *Qual Quant* 47(4):2025–2047
- Kühne S (2018) From strangers to acquaintances? Interviewer continuity and socially desirable responses in panel surveys. *Surv Res Methods* 12(2):121–146
- Leopold T, Skopek J (2015) Convergence or continuity? The gender gap in household labor after retirement. *J Marriage Fam* 77(4):819–832
- Lippmann Q, Georgieff A, Senik C (2020) Undoing gender with institutions: lessons from the German division and reunification. *Econ J* 130(629):1445–1470
- Lipps O, Lutz G (2017) Gender of interviewer effects in a multi-topic centralized CATI panel survey. *Methods Data Anal* 11(1):67–86
- Liu M, Stainback K (2013) Interviewer gender effects on survey responses to marriage-related questions. *Pub Opin Q* 77(2):606–618
- Lueptow LB, Moser SL, Pendleton BF (1990) Gender and response effects in telephone interviews about gender characteristics. *Sex Roles* 22(1–2):29–42
- Murray-Close M, Heggeness ML (2019) Manning up and womaning down: how husbands and wives report earnings when she earns more. Opportunity and Inclusive Growth Institute Working Papers, 28
- Paulhus DL (1991) Measurement and control of response bias. In: Robinson JP, Shaver PR, Wrightsman LS (eds) Measures of personality and social psychological attitudes. Academic Press, San Diego, pp 17–59
- Press JE, Townsley E (1998) Wives’ and husbands’ housework reporting: gender class and social desirability. *Gend Soc* 12(2):188–218
- Roth A, Slotwinski M (2020) Gender norms and income misreporting within households. ZEW Discussion Paper, No. 20-001
- Schnell R, Kreuter F (2005) Separating interviewer and sampling-point effects. *J Off Stat* 21(3):389–410
- Schober PS (2013) Maternal labor market return and domestic work after childbirth in Britain and Germany. *Community Work Fam* 16(3):307–326
- Schober PS, Zoch G (2019) Change in the gender division of domestic work after mothers or fathers took leave: exploring alternative explanations. *Eur Soc* 21(1):158–180
- Sullivan O (2010) Changing differences by educational attainment in fathers’ domestic labour and child care. *Sociology* 44(4):716–733
- Sullivan O (2011) An end to gender display through the performance of housework? A review and reassessment of the quantitative literature using insights from the qualitative literature. *J Fam Theory Rev* 3(1):1–13
- Sullivan O (2013) What do we learn about gender by analyzing housework separately from child care? Some considerations from time-use evidence. *J Fam Theory Rev* 5(2):72–84
- Tourangeau R, Rips LJ, Rasinski KA (2000) The psychology of survey response. Cambridge University Press, Cambridge
- Tourangeau R, Yan T (2007) Sensitive questions in surveys. *Psychol Bull* 133(5):859–883
- Warner RL (1986) Alternative strategies for measuring household division of labor. *J Fam Issues* 7(2):179–195
- West BT, Blom AG (2017) Explaining interviewer effects: a research synthesis. *J Surv Stat Methodol* 5(2):175–211
- West C, Zimmerman DH (1987) Doing gender. *Gend Soc* 1(2):125–151

Zoch G (2021) Gender-of-interviewer effects in self-reported gender ideologies: evidence based on interviewer change in a panel survey. *Int J Public Opin Res* 33(3):626–636

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.