Why Should Women Take Care of the Housework?

Experimental Evidence Concerning Gender-Specific Norms for the Division of Paid and Unpaid Work Among Couples

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Sozialwissenschaftlichen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von

Sabine Manuela Düval

2023

Erstgutachterin:	Prof. Dr. Katrin Auspurg Ludwig-Maximilians-Universität München
Zweitgutachter:	Prof. Dr. Josef Brüderl Ludwig-Maximilians-Universität München
Tag der mündlichen Prüfung:	27.02.2023

Danke!

Ich möchte mich bei allen Personen bedanken, die einen maßgeblichen Anteil an der Realisierung der vorliegenden Arbeit haben.

Mein herzlicher Dank gilt zunächst der Erstbetreuerin meiner Arbeit, Prof. Dr. Katrin Auspurg, für ihr wertvolles Feedback, ihre Anregungen und ihre Unterstützung. Auch bei meinem Zweitbetreuer, Prof. Dr. Josef Brüderl, möchte ich mich für seine Ideen und Perspektiven sehr bedanken. Prof. Dr. Sabine Walper danke ich für die Übernahme der Rolle als Drittprüferin meiner Disputation.

Einen wichtigen Anteil an dieser Arbeit haben meine (ehemaligen) Kolleg:innen an den Lehrbereichen Auspurg und Brüderl. Danke euch für die angenehme Arbeitsatmosphäre und für die vielen hilfreichen Hinweise und Ideen bei diversen Kolloquien und Writing Workshops. Ganz besonders möchte ich mich bei Werner Fröhlich, Christian Ganser, Renate Lorenz, Bettina Pettinger und Fabian Thiel bedanken. Eure fachliche und ganz besonders auch eure persönliche Unterstützung haben mir durch manch schwierige Phase geholfen. Danke auch dir, Maggi, für die freundschaftliche Standleitung nach Barcelona.

Eine besondere Erwähnung gilt meinen Eltern, die mich beständig begleiten und mich in jeglicher Hinsicht unterstützen. Danke für euer Vertrauen in meinen Weg. Und schließlich möchte ich mich bei dir bedanken, Daniel, dass du in jeder Phase meiner Promotion an meiner Seite gestanden bist, mich motiviert und unterstützt hast und mir die (teilweise dringend notwendige) Ablenkung geboten hast.

München, April 2023 Sabine Düval

Zusammenfassung

Diese Dissertation liefert einen Beitrag zur Erforschung von geschlechtsspezifischen Normen im Kontext der Haus- und Erwerbsarbeitsteilung von heterosexuellen Paaren. In allen westlichen Gesellschaften übernehmen Frauen mehr Hausarbeit und Kinderbetreuung als Männer. Häufig wird argumentiert, dass Geschlechternormen die Arbeitsteilung auf der individuellen Ebene dahingehend beeinflussen, dass Individuen ihre (Haus-)Arbeiten geschlechtsspezifisch aufteilen.

Ich kritisiere an dieser Argumentation, dass der Einfluss von Geschlechternormen in der bestehenden Literatur nur oberflächlich erklärt wird. Mögliche Mechanismen, die den Zusammenhang erklären könnten, werden kaum diskutiert. Vielversprechend könnte die Erklärung sein, dass Paare normkonform handeln, um Kosten (z.B. soziale Sanktionen) zu vermeiden, die durch eine Abweichung von der Geschlechternorm entstehen. Nur wenn die Kosten höher sind als mögliche Gewinne durch eine egalitäre Arbeitsteilung, wäre ein Effekt von Geschlechternormen zu erwarten. Das setzt eine weitverbreitete Geschlechternorm voraus, deren Nichtbefolgung sanktioniert wird.

Weiter argumentiere ich, dass die Literatur zu Geschlechternormen ein wichtiges Merkmal sozialer Normen weitgehend ignoriert – nämlich ihre Bedingtheit (d.h., dass sie kontextabhängig angewandt werden). Das Nichtbeachten dieser Bedingtheit hat insbesondere Auswirkungen auf die Messung von Geschlechternormen. In der Regel werden Geschlechternormen mit Item-Fragen gemessen. Befragte bewerten kurze Aussagen zur Zuständigkeit von Frauen und Männern in Hausarbeit, Kinderbetreuung und Erwerbsarbeit. Die Items erhalten keine tiefergehenden Informationen zur sozialen Situation, in der sich die Frauen und Männer befinden. Bei einer bedingt angewandten Geschlechternorm ist diese Information aber notwendig, um die Situation zu bewerten. Ich vermute daher, dass Befragte diese fehlenden Informationen eigenständig unterstellen. Das würde die Messung von Geschlechternormen verzerren.

Deshalb untersuche ich, ob Geschlechternormen bedingt angewandt werden. Genauer gesagt beleuchte ich, unter welchen Bedingungen Individuen (normativ) erwarten, dass Frauen unbezahlte Arbeiten übernehmen sollen. Um diese Forschungslücke zu schließen, habe ich ein faktorielles Survey-Experiment zur Aufteilung von bezahlter und unbezahlter Arbeit von Paaren erstellt. In Vignetten bzw. in ihren Dimensionen wird der Haus- und Erwerbsarbeitsanteil einer fiktiven (männlichen oder weiblichen) Vignettenperson beschrieben. Diese Information wird zufällig variiert.

4 624 Befragte im Alter von 23 bis 47 Jahren haben 13 703 Vignetten bewertet, die 2017/18 im Beziehungs- und Familienpanel (pairfam) erhoben wurden. Die Befragten sollten die Angemessenheit der Hausarbeitsteilung oder der Aufteilung der gesamten Arbeit (Hausarbeit, Erwerbsarbeit und Kinderbetreuung) bewerten. Zur Beantwortung meiner Forschungsfrage habe ich beide Experimentalbedingungen in zwei verschiedenen Studien analysiert.

Die erste Studie konzentriert sich auf die Bedingtheit von Geschlechternormen. Konkret untersuche ich, ob es einen direkten Effekt des Geschlechts der Vignettenperson auf die Angemessenheit des Hausarbeitsanteils gibt. Dafür entkopple ich den Geschlechtereffekt von anderen (mediierenden) Faktoren. Ich finde Belege für meine Annahme, dass Geschlechternormen bedingt angewendet werden. Tatsächlich wollen Befragte nur dann, dass Frauen für die Hausarbeit zuständig sind, wenn sie weniger Erwerbsarbeit leisten als ihre männlichen Partner oder wenn sie weniger verdienen. Das kann als Nachweis für eine Equity-Norm gesehen werden. Ich finde keine Belege für die Existenz einer traditionellen Geschlechternorm.

Dieses Ergebnis wird durch meine zweite Studie bestätigt. In dieser untersuche ich, ob sich Männer und Frauen in ihren Geschlechternormen unterscheiden. Besonderes Augenmerk lege ich dabei auf die Mehrdimensionalität von Geschlechternormen. Ich finde keine Evidenz dafür, dass sich männliche und weibliche Befragte in ihren Bewertungen unterscheiden. Beide verwenden Equity- anstelle von Geschlechternormen.

Abschließend weise ich auf eine mögliche Limitation hin: Befragte sind einer hohen kognitiven Belastung ausgesetzt, wenn sie ein faktorielles Survey-Experiment beantworten. Es ist möglich, dass ältere oder niedriger gebildete Befragte nicht die gesamte in den Vignetten dargestellte Information erfassen. Ebenso könnten sehr schnell Antwortende Teile der Information überlesen. Das hätte einen Effekt der Dimensionsreihenfolge zur Folge.

Um Effekte der Dimensionsreihenfolge zu untersuchen, verwende ich ein zweites faktorielles Survey-Experiment, das in eine Bevölkerungsumfrage der Stadt Konstanz eingebunden war. Die Befragten wurden zufällig in vier Gruppen mit unterschiedlicher Dimensionsreihenfolge aufgeteilt. Das erlaubt zu untersuchen, ob die Reihenfolge das Urteil beeinflusst. Zusätzlich analysiere ich verstärkende Effekte durch Alter, Bildung und Antwortgeschwindigkeit. Ich finde keine deutlichen Effekte der Dimensionsreihenfolge. Es zeigt sich nur, dass sehr schnell Antwortende etwas zu Recency-Effekten neigen, d.h. dass ihre Bewertungen überproportional stark auf der letzten Vignettendimension basieren. Diese kleinen Effekte dürften meine Schlussfolgerungen zur Geschlechternorm nicht beeinträchtigen.

Zusammengefasst findet diese Dissertation Evidenz für eine Equity-Norm im Kontext der Arbeitsteilung von Paaren. Unter Berücksichtigung, dass soziale Normen bedingt angewandt werden, gibt es keine Hinweise auf geschlechtsspezifische Normen. Folglich ist es unwahrscheinlich, dass Individuen oder Paare mit starken sozialen Sanktionen zu rechnen haben, wenn sie ihre Haus- und Erwerbsarbeit nicht auf Grundlage des Geschlechts aufteilen. Dieses Ergebnis hinterfragt die weit verbreitete Annahme, dass die ungleiche Aufteilung von bezahlter und unbezahlter Arbeit zwischen Männern und Frauen (zumindest in Teilen) auf traditionelle Geschlechternormen zurückzuführen ist.

Abstract

This dissertation advances research on gender-specific norms underlying the division of paid and unpaid work among couples. In all Western societies, women continue to be responsible for more of the housework and child care than men. Gender norms are often cited as one of the explanations for this gender difference at the individual level. In other words, it is assumed that individuals who have internalized traditional gender norms opt for a gendered division of work.

I argue that the theoretical discussion on the effect of gender norms remains rather superficial in existing literature. There is little discussion of an actual mechanism that explains this effect. One promising explanation is that couples behave in a norm-conforming manner to avoid costs that could arise from a possible deviation, such as social sanctions due to deviations from an (internalized) social norm. These costs must trump possible gains from an egalitarian division of work to traditionalize couples' work division, which requires the existence of a strong (traditional) gender norm.

Moreover, I note that existing research on gender norms largely fails to discuss an important feature of social norms, namely their conditionality. The neglect of a possible conditionality of (social) gender norms has far-reaching implications for this area of research, and is particularly relevant for measuring (internalized) gender norms. Gender norms are typically measured using item questions. Respondents agree or disagree with general statements about men's and women's responsibilities for housework, child care, and paid work. These item questions, however, lack information on the social context of these hypothetical men and women. I argue that when gender norms are conditionally applied, respondents are likely to impute this lacking information based on the social context. If so, the measurement of gender norms would be biased.

For this reason, I probe whether gender norms are applied conditionally and if so, under which conditions individuals feel that women should be responsible for unpaid work. To close this research gap, I conducted a factorial survey experiment on the division of paid and unpaid work among couples. Respondents received short descriptions (i.e., vignettes) of the division of work among fictional couples. Each vignette contained information on the gender of the vignette person and their work-family arrangement (e.g., time spent in housework and paid work and relative income), and these vignette dimensions were varied randomly.

A total of 4,624 respondents aged 23 to 47 took part in the factorial survey experiment on the division of paid and unpaid work implemented in the German Family Panel (pairfam) in 2017/18. In this experiment, a total of 13,703 vignette evaluations were collected. By random assignment respondents were attributed to evaluate the appropriateness of the vignette person's share of the housework only or of the share of the total workload (i.e., housework, paid work, and child care). To answer my research questions, I analyzed both experiments in two different studies.

The first study focuses on probing the conditionality of gender norms. In short, I disentangle the direct effect of the vignette person's gender from other (mediating) effects on respondents' judgement of an appropriate housework division. There is much support for my assumption that gender norms are applied conditionally. In fact, only in the case that women work relatively less hours in paid work than their male partners or if they earn relatively less, do respondents hold them responsible for housework. This is evidence for the existence of an equity norm in the context of couples' housework division; I find no support for the existence of a gender norm.

This finding is further supported by my second study. Here, the focus is on describing whether men and women differ in their gender norms. I use the experimental setup to capture different aspects of the multidimensional concept of gender norms. Both groups apply equity rather than gender norms to rate the appropriate division of the total workload—men and women do not differ here.

Finally, I acknowledge a possible limitation of the two studies, namely the cognitive load placed on respondents when evaluating a factorial survey experiment. As an example, I examine whether the order in which the vignette dimensions are presented to respondents has an impact on their evaluations. It is possible that respondents are not able to grasp all the presented information due to cognitive overload. Likewise, very fast respondents might not consider all the information. As a result, dimension order effects are possible.

To examine effects of the dimension order, I use a factorial survey experiment included in a population survey of the German city Konstanz. The respondents were randomly attributed to four different experimental groups, each with a different dimension order. To detect possible dimension order effects, I analyzed different respondent groups across age, education, and response speed. In sum, I find no strong support for a dimension order effect. There is only small evidence that fast responders are more prone to so-called recency effects (i.e., their evaluations are disproportionately based on the last vignette dimension). However, these minor effects should not seriously affect the main conclusions of this dissertation.

To conclude, this dissertation finds widespread support for a division of paid and unpaid work based on equity norms. Considering that social norms are applied conditionally, there is no evidence for gender-specific norms underlying the division of work among couples. It is therefore unlikely that individuals or couples would receive strong social sanctions if they base their division of work on relative income and/or on hours spent in paid work rather than on gender. These results challenge the common assumption that the unequal distribution of paid and unpaid work between men and women is (at least partly) due to traditional gender norms.

Contents

List of Tables

List of Figures

1

Conceptual Framework: Why Should Women Take Care of the Housework?

(Sabine Düval)

1 Conceptual Framework: Why Should Women Take Care of the Housework?

1.1 Introduction

1.1.1 The German Context: Couples' Division of Paid and Unpaid Work

In Western countries such as Germany, gender differences persist in key dimensions of social inequality like occupation and income. For example, the gender gap in hourly wages (the gender wage gap) in Germany was roughly 18% in 2021 (German Federal Statistical Office, 2022b). In 2019, the gender gap in pensions was even higher with about 36% (Eurostat, 2022). But why exactly does gender, an ascriptive characteristic, influence individuals' income and occupation and, consequently, their physical and social well-being?

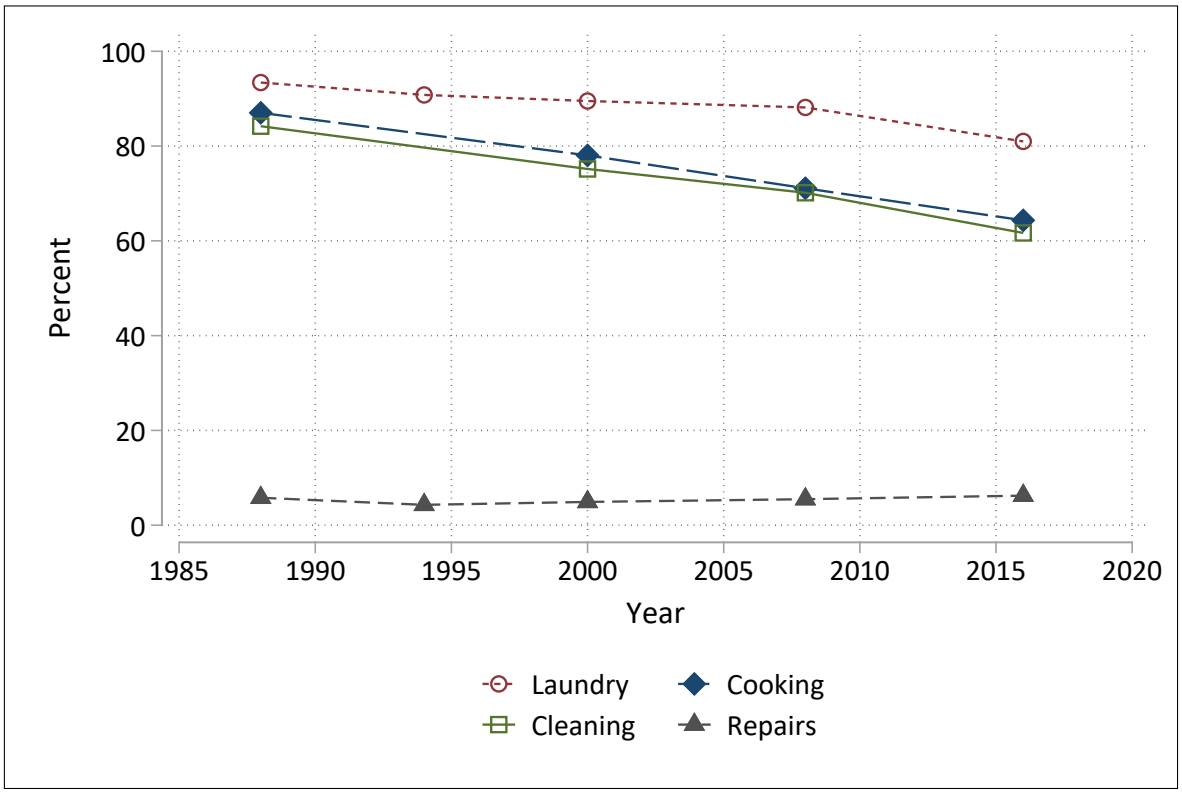


Figure 1.1. Percentage of Women Who Are Always or Most Often Responsible for Selected Household Tasks Over Time (1988-2016)

Source: ALLBUS 1982-2018 (GESIS - Leibniz-Institut für Sozialwissenschaften, 2021). Weighted due to an over-representation of East German households. Own calculations.

The work-family conflict for women has often been discussed as one reason for the gender-specific differences in occupation and income (see, e.g., Kaufman and Taniguchi, 2019; McGinnity and McManus, 2007; Trappe et al., 2015, and many more). Women spend significantly more time each day doing unpaid work than men. For Germany,

Figure 1.1 shows the percentage of women in heterosexual partnerships who are always or most often responsible for selected household tasks (i.e., laundry, cooking, cleaning, and repairs). In 1988, roughly 90% of women were always or mostly responsible for laundry, and 80% were responsible for cooking or cleaning. Even though the gender gap in housework has decreased over the last decades, 80% of women were still always or most often responsible for laundry in 2016. The same pattern (albeit at a slightly lower level) can be observed for other time-consuming routine household chores such as cooking and cleaning, each of which was primarily the responsibility of about 60% of women in 2016. Typically, female housework tasks must be completed daily or at least several times per week, whereas men are typically responsible for tasks that only need to be done at irregular intervals such as repairs or maintenance (see, e.g., Blair and Lichter, 1991; Coltrane, 2000; Tai and Treas, 2012). The existence of a gender care gap, which manifests itself in women providing about twice as much child care as men during the week (for Germany, see, e.g., Walper and Lien, 2018), also points to women's one-side responsibility for unpaid work. As a result, this responsibility for the majority of unpaid work creates a conflict with women's careers: they often must reduce time spent in paid employment when unpaid tasks require their attention (see, e.g., Grunow et al., 2007, 2012; Gupta et al., 2021; Kühhirt, 2012; Nitsche and Grunow, 2016; Solera and Mencarini, 2018; Sullivan, 2021).

1.1.2 Research Question

Many researchers have attempted to understand the mechanism by which gender affects the proportion of paid and unpaid work for which an individual is responsible. In short, the explanations boil down to two different mechanisms as the drivers for gender differences in paid and unpaid work: gender norms and economic resources. Both approaches have been widely studied, confirmed, but also rejected in the past decades (for a recent overview, see, e.g., Grunow, 2019). As a consequence, it is still somewhat unclear as to why women are responsible for the lion's share of domestic labor.

This dissertation does not join the multitude of essays that attempt to find determinants for the division of paid and unpaid work among couples. Instead, I ask whether gender norms are (still) at all suitable as an explanation for the unequal distribution of paid and unpaid work between men and women. From a theoretical perspective, I discuss whether the mechanisms behind gender norm approaches are useful in explaining the link between gender and shares of paid and unpaid work. Empirically, I aim to uncover whether and why it is believed that women should be responsible for child care and housework: Is there (still) a gender-specific social norm that expects women to do unpaid domestic labor?

To answer these questions, I will trace the two (prominent) theoretical approaches that focus on gender as the main predictor for couples' division of paid and unpaid

labor: the gender norm approach and the gender deviance neutralization approach. In short, the gender norm approach specifically argues for the existence of gender-specific norms that drive individuals' behavior (see e.g., Behr et al., 2013; Bittman et al., 2003; Brines, 1993, 1994; Carlson and Lynch, 2013; Davis and Greenstein, 2009). These gender-specific norms that expect women to be responsible for housework and child care and men to not be responsible for these areas are internalized by both men and women and influence their expectations of an appropriate division of labor (Bittman et al., 2003). It is argued that couples who comply with these gender norms choose a traditional division of paid and unpaid work (i.e., a male breadwinner and a female homemaker who additionally takes care of the couple's children). Egalitarian couples would then divide paid and unpaid work regardless of gender.[1] In contrast, the gender deviance neutralization approach is based on an argument by West and Zimmerman (1987) that individuals "do gender", meaning that they exhibit gender-specific behaviors in order to portray their gender.[2] According to this approach, women who deviate from their normative roles as homemakers (e.g., by earning more than their male partners) are pushed to compensate for this deviation by taking over most (if not all) of the housework and child care tasks (Brines, 1993, 1994). In the same vein, men who deviate from their breadwinner role are expected to decrease their contributions to domestic tasks.

I argue that the mechanisms behind both of these approaches are vague, at best. A rational choice-based approach, for example by Akerlof and Kranton (2000), would be more helpful to uncover the true mechanism through which women are expected to do more domestic tasks. In short, they argue that individuals follow gender norms to increase utility or rather to avoid social sanctions that arise from not complying (Akerlof and Kranton, 2000).

Other than the existence of sanctions in cases of non-compliance, another characteristic of social norms is their conditionality (see, e.g., Axelrod, 1986; Elster, 2018). As a result, social gender norms can also be expected to be applied conditionally. In other words, depending on the social situation, gender norms likely lead to different behavioral expectations. For example, if it is generally assumed that mothers who contribute less to household income than their male partners should be responsible for housework and child care, this does not necessarily mean that this is also expected in situations in which women contribute the same amount of income or when the couple has no children. This aspect has only been implicitly discussed in the previous literature, if at all. Thus, it is not only unclear if gender norms are (still) applied in the context of

[1] Recent research argues that gender norms should not be interpreted on this unidimensional scale from "egalitarian" to "traditional" (see Grunow et al., 2018), however, the general notion that norms or gender role attitudes affect behavior is still the same.

[2] This approach is also known as gender deviance compensation, gender display or gender construction (e.g., Bittman et al., 2003; Brines, 1993, 1994; England, 2011; Sullivan, 2011).

division of work, but also whether they are applied unconditionally or conditionally; and if they are conditionally applied, the conditions under which women should do the unpaid work are not clear.

In addition, I probe whether different sub-groups apply gender norms differently. For example, men profit from a social norm that does not expect them to do housework. In result, they might apply gender norms unconditionally and always believe that women should be responsible for unpaid work. Women, however, might apply gender norms conditionally, as they benefit when the social situation is taken into account.

If social gender norms are, in fact, applied conditionally, I argue that they cannot be validly measured with standard item questions in which respondents are asked, for example, to express their views on women or mothers being employed. Many limitations of these questions have already been emphasized by other authors (for a recent overview, see Walter, 2018b). In this dissertation, I introduce an additional shortcoming of standard item questions. In short, item questions on gender norms do not provide enough information on the social situation that is to be judged. Respondents must agree or disagree with abstract and short normative statements about men and women. However, in order to decide which (conditional) norm is to be applied here and if the described men and women act according to this norm, information on the social situation is required. I therefore argue that the effect of gender norms could be overestimated if the abstract item wording triggers associations with traditional divisions of work among respondents. This puts the validity of empirical results on the influence of gender norms in question when measured using standard item questions.

To answer all these questions, I use a factorial survey experiment conducted in 2017/18 as part of the large-scale German Family Panel pairfam.[3] In total, 4,624 German residents aged 23 to 47 evaluated 13,703 vignette questions about hypothetical couples' work-family arrangements. The gender of the described vignette persons (i.e., the actors described in the factorial survey experiment) was randomly varied. In addition, the amount of information and the specific content concerning vignette persons' work-family situation (e.g., both partners' employment, shares of unpaid work, and family status) was also varied randomly. This allows for an understanding of whether a gender norm is present that expects men and women to differ in their contributions to paid and unpaid work under any and all circumstances, or whether gender norms are instead applied conditionally, resulting in different behavioral expectations depending on the social situation (e.g., women's employment situation or the existence of a child).

[3] Analyses are based on data of the German Family Panel (pairfam), release 10.0 (see Brüderl et al., 2019). A detailed description of the study can be found in Huinink et al. (2011). This dissertation uses data from the German Family Panel pairfam, coordinated by Josef Brüderl, Sonja Drobnič, Karsten Hank, Johannes Huinink, Bernhard Nauck, Franz J. Neyer, and Sabine Walper. From 2004 to 2022 pairfam was funded as priority program and long-term project by the German Research Foundation (DFG).

Factorial survey experiments are well-suited to measuring social norms (see, e.g., Jasso and Opp, 1997; Opp, 2002, 2020). However, answering a factorial survey module is cognitively challenging (see Auspurg and Hinz, 2015), and learning and/or fatigue effects might bias the results. To address possible method effects inherent to factorial survey experiments, I examine one exemplary constraint: whether the order in which situational information (i.e., dimensions) is presented in the factorial survey experiment affects the results of the experiment.

In this chapter, I will justify the research question of this dissertation and answer it based on the joint results of three research articles. Each of the three stand-alone research articles then follow as separate chapters.

1.2 State of Research on Gender Norms

1.2.1 Prevalence of Gender Norms in Germany

As gender norms are a latent construct, they cannot be measured by direct questions (see Davis and Greenstein, 2009; McHugh and Frieze, 1997). Traditionally, population surveys such as the General Social Survey in the USA, the German General Social Survey, the International Social Survey Programme, the Word Value Survey, and many more have included scales into their respective questionnaires as a proxy for gender norms (see Davis and Greenstein, 2009, for more examples and detailed information on the gender norm scales in the surveys mentioned, see e.g., Beere, 1990, Kroska, 2007, Walter, 2018a). Typically, these scales consist of a number of item questions with three to seven response options that vary from strong agreement to strong disagreement (see Kroska, 2007). Individuals' ratings of these item batteries are then interpreted as gender role attitudes,[4] based on which individuals are then either characterized as having egalitarian or traditional gender role attitudes.[5] For example, agreeing with the statement "Men should participate in housework to the same extent as women" (see Table 1.1, Item 1) is understood as an egalitarian gender role attitude, while agreeing with "Women should be more concerned about their family than about their career" (see Table 1.1, Item 2) is interpreted as a traditional gender role attitude. It

[4] A variety of phrases is used to describe individuals' support for a division of paid and unpaid work that is based on separate spheres for each gender. Further phrases are gender ideology, gender role, attitudes about gender, gender-related attitudes, and gender egalitarianism (see Davis and Greenstein, 2009). In this dissertation, the focus is on gender-specific norms. I use the terms "gender norms", "social gender norms" or "gender-specific norms" when referring to the concept of a social norm that is based on gender. When referring to individuals' internalizations of this social gender norm, I use "internalized gender norms" and "gender role attitudes"; the first to relate to the literature on social norms, the latter to consider that the standard way of measuring internalized gender norms is with attitude scales.

[5] Newer research argues why this unidimensional measurement of gender role attitudes is outdated and why gender norms should be interpreted as multidimensional constructs (see Grunow et al., 2018).

is (usually implicitly) assumed that the prevailing gender norm can be derived from these individual attitudes. Therefore, out of the cumulative evaluations of the gender role item questions, gender norms of societies or subgroups can be inferred. Table 1.1 shows the gender role items included in wave 10 (2017/18) of the German Family Panel pairfam.

Table 1.1. Classical Item Questions on Gender Norms in the German Family Panel Pairfam, Wave 10 (2017/18)

Item 1: Men should participate in housework to the same extent as women.						
Item 2: Women should be more concerned about their family than about their career.						
Item 3: A child under 6 will suffer from having a working mother.						
Item 4: Children often suffer because their fathers spend too much time at work.						
Disagree completely	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	Agree completely

Over the past forty years, traditional gender norms towards the division of paid and unpaid work between men and women have been on the decline (for Germany, see e.g., Eyerund and Orth, 2019; Grunow et al., 2018); at least when measured with the standard item questions. Figure 1.2 shows that in the 1980s, roughly 70% of Germans expressed traditional gender role attitudes by agreeing with the statement "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." Over the years, this agreement declined to between 20% and 30% in 2016.[6] In all years, men expressed more traditional attitudes than women (e.g., 28% vs. 20% agreement in 2016).[7]

1.2.2 Research on Division of Paid and Unpaid Work

Since the classic studies in the 1960s and 1970s (e.g., Bernard, 1972; Blood and Wolfe, 1960) more research has been published to explain why women do more housework than men. Gender norms have often been found to influence both men's and women's shares of housework in child care in Germany, the United States, Great Britain, and many other OECD countries. At the same time, research has found also support for other relevant predictors of a couples' work division, such as partners' relative earnings and time spent in paid work.[8] For overviews, see Coltrane (2000) for the 1990s, Lachance-

[6] Note that a decline in individuals' traditional attitudes cannot only be attributed to the German reunification in 1990. Although more (former) East Germans expressed egalitarian attitudes than (former) West Germans, a decline in agreement with the item questions can also be seen among (former) West Germans (see Figure 1.A1 in the Appendix).

[7] For Germany, see also Arránz Becker (2013) and Horne and Johnson (2018), for the US, see Amato and Booth (1995), and Meagher and Shu (2019).

[8] These alternative (economic) predictors are discussed in resource theory, family economics, and the time availability approach.

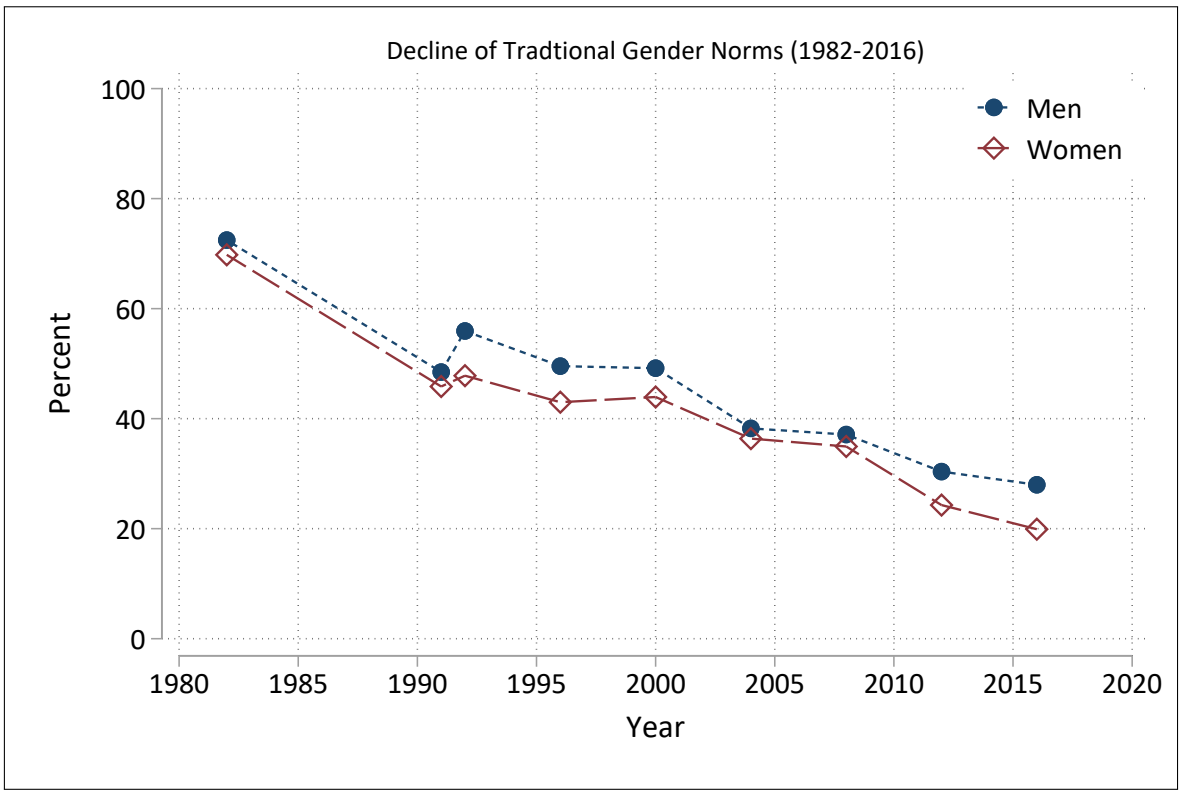


Figure 1.2. Gender Difference in Decline of Traditional Gender Norms in Germany (1982-2016).

Source: ALLBUS 1982-2018 (GESIS - Leibniz-Institut für Sozialwissenschaften, 2021). Weighted due to an over-representation of East German households. For 2012 and 2016 only split A was used. Own calculations.

Grzela and Bouchard (2010) for the 2000s, and Grunow (2019) for the 2010s.

Another relevant finding concerns the interplay of gender norms and relative resources. There is an agreement in the literature that gender role attitudes moderate the effect of other factors such as economic resources (Coltrane, 2000; Davis and Greenstein, 2009; Lachance-Grzela and Bouchard, 2010). For individuals with egalitarian gender role attitudes, the effect of resources (e.g., income) is greater than for individuals with traditional gender role attitudes. Other studies have found that men who depend financially on their female partners do less housework than their partners, which is interpreted as support for gender deviance neutralization (e.g., Bittman et al., 2003; Brines, 1994).

Overall, there is support that gender norms, resources, and time availability at least partly explain the division of paid and unpaid work within couples. Nevertheless, the question remains as to why women do more housework than men. In the next section, I introduce some of the research that suggest there is still much (theoretical and methodical) work to be done to uncover the exact mechanism for couples' unequal division of work.

For example, Sullivan (2011) and England (2011) point out the importance of not

over-interpreting existing evidence for gender deviance neutralization, as the subgroup of men who earn less than their female partners is highly selective and consists mainly of men who were long-term unemployed and men belonging to the lowest of income percentiles. In addition, evidence for gender deviance neutralization has been partially based on incorrectly specified equations (see England, 2011; Sullivan, 2011).

Killewald and Gough (2010) contested previous studies who found support for gender deviance neutralization by examining the effect of outsourcing housework on women's housework time. They found that women's housework time decreases more if low-earning women experience an increase in earnings compared to an increase in earnings for high-earning women. They explain this result by stating that high-earning women have already stopped doing housework that is easy to outsource, whereas for low-earning women, an increase in earnings leads to the outsourcing of housework responsibilities. They argue that previous studies that found support for gender deviance neutralization did not account for this non-linear relationship; when including information on outsourcing of housework, the authors found no evidence for gender deviance neutralization.

Carlson and Lynch (2013) even questioned the assumed one-sided effect of attitudes on behavior and found support for their assumption of a reciprocal relationship in which a couple's division of work also influences their gender role attitudes. This result is highly relevant, as it questions existing research on the effect of gender role attitudes – at least when analyzed with cross-sectional approaches.

Results are also dependent on the methods used to operationalize the research question. For example, Schneider (2012) found that men and women compensate gender-atypical employment by showing gender-typical housework performance. Using the same data from the National Survey of Families and the Household and the American Time Use Survey, McClintock (2017) did not find support for the hypothesis. Instead, she found that men and women who work in gender typical occupations perform gender typical housework, while men and women in gender atypical occupations perform gender atypical housework tasks. In contrast to Schneider (2012), who used quadratic terms to model nonlinearity, McClintock (2017) specified models that included categorical dummy variables to allow for a closer look at the non-linear association.

In addition, the measurement of relevant constructs has also been contested in existing literature. Yavorsky et al. (2015) showed that results are highly dependent on the measurement instrument used by comparing men's and women's time in housework measured by a conventional survey and by a time diary: If researchers had only relied on conventional survey data, they would have underestimated the gender gap in housework time after the birth of the first child. Especially men overestimated their time spent in housework in the conventional survey. Similar conclusions can be drawn from Klein and Kühhirt (2010), who found that men in particular are prone to method

effects (e.g., due to social desirability) when reporting their time spent doing housework and therefore overestimate it. Also, the measurement of gender role attitudes via item questions has been criticized in the past (for overviews, see Walter, 2018b). However, to date, most analyses using the criticized item questions only discuss this critique in a side note, if at all. Interestingly, some studies find support for an effect of gender norms without including a measurement of gender norms in their analyses (e.g., Grunow et al., 2012; Kühhirt, 2012)—they simply assume that a residual effect must be explained by gender norms. However, a valid measurement of internalized gender norms (or gender role attitudes) is elementary to a discussion of a possible effect of gender norms on the division of paid and unpaid work.

In sum, there is much evidence that previous research on men's and women's shares of work is not only highly dependent on the exact research question, but also on the methods, measurement instruments, and control variables (e.g., which mediator or confounder variables are included). Of course, research has progressed over the past decades. For example, fixed-effects regressions have been used in place of simple cross-sectional analyses to control for unobserved unit heterogeneity (for examples using German data, see Cordero-Coma and Esping-Andersen, 2018; Dechant et al., 2014; Leopold and Skopek, 2015; Procher et al., 2018). Other analyses have focused on country comparisons to investigate differences in institutional contexts, for example between welfare regimes, work-family policies, and gender equality. Landmark studies such as Fuwa (2004) and Fuwa and Cohen (2007) both use data from the International Social Survey Programme and come to the conclusion that in countries that are more egalitarian on the macro-level (e.g., concerning gender equality in economic and political power, female labor-force participation, parental leave policies, or child care services), housework is divided more equally between men and women (see also Dotti Sani, 2014; Heisig, 2011; Thébaud, 2010). However, these main results are contested by Gangl and Ziefle (2015), who found (at least for Germany) that an increase in parental leave resulted in a decline in mothers' work commitment.

What is evident from these few examples is that most existing research on gender inequalities in the division of (house)work tests the same hypotheses over and over, each derived from existing theory. Theoretical conclusions are then drawn from the results that either support or reject the theories (for a similar critique, see England, 2011). However, a deeper exploration of the existing theoretical concepts is lacking. Besbris and Khan (2017, p. 147) argue that (especially) while a concept may have been highly used and reformulated for decades, this does not mean that the concept itself is immune to criticism. Instead, it is necessary to thoroughly examine the reasoning underlying a theoretical concept and unravel the mechanism commonly used to explain a (causal) relationship of interest. In the next section, I therefore elaborate on the theoretical concept behind the gender norm approach and look into various mechanisms used to

explain the relationship between (gender-specific) norms and the division of work.

1.3 Theoretical Framework

1.3.1 On the Conditionality of Gender Norms

At first glance, the concept of gender norms seems intuitive; a (universal?) norm that governs men's and women's behavior, not only in regard to their intra-family work division, but also in regard to other areas of life (e.g., Davis and Greenstein, 2009). Parsons (1955) observed different expectations for men and women based on their gender, with a woman being held responsible for "the internal affairs of the family, as wife, mother and manager of the household [, while a man is] primarily anchored in the occupational world, in his job and through it by his status-giving and income-earning functions of the family" (Parsons, 1955, p. 14f.). He also interprets gender as the primary reason for a gender difference in shares of paid and unpaid work (Parsons, 1959, p. 264) and implies that it is unlikely that these roles would be reversed, even if the female partner were to work outside the home (Parsons, 1955, p. 15). This can be interpreted as the existence of an unconditional gender norm that drives men's and women's behavior. Furthermore, by arguing that men and women can only love a member of the opposite sex who adheres to the norm set by his or her gender, Parsons (1955, p. 22) (implicitly) argues that these unconditional gender norms are internalized by both men and women.[9] He argues that this internalization of gender norms is rooted in both socialization and in biological gender differences in childbearing and breastfeeding (Parsons, 1955, 1959), which leads to gender differences in occupation and income (Parsons, 1959).

Not only Parsons argues that gender norms govern individuals' behavior independent of the situational context. Similarly, the prominent argument that "gender trumps money" predicts that it is (mainly) gender that predicts who is responsible for a heterosexual couple's (routine) housework, even if the female partner has a similar or even greater labor market power than the male partner (see Bittman et al., 2003; Brines, 1993, 1994; Hochschild and Machung, 1989). In this case, gender norms would hold even when the male breadwinner stereotype does not apply. This points to an unconditional gender norm for women to be mainly responsible for housework.

Usually, social norms only hold under certain conditions, meaning they are not applied unconditionally (see, e.g., Beck and Opp, 2001; Bicchieri, 2008, 2010; Bicchieri and Chavez, 2010; Diefenbach and Opp, 2007; Elster, 2018; Kahnemann et al., 1986, and many more). Instead, they are interpreted as being "rather local and context dependent" (Bicchieri, 2008, p. 229). Even if a social norm is formulated unconditionally,

[9] It also becomes clear here that Parsons (1955) bases his argumentation exclusively on heterosexual men and women who identify with their gender assigned at birth.

this does not imply that the norm holds under any circumstance (e.g., Diefenbach and Opp, 2007, p. 489). For example, Hechter and Opp (2001) and Diefenbach and Opp (2007) argue that even the norm of truthfulness is only applied conditionally.[10] Jasso and Opp (1997, p. 948) further argue that even for the strict norm against murder, there are conditions under which killing another human is allowed, for example in situations of war or in self-defense. Therefore, what is appropriate or fair depends on context-dependent normative expectations of what one ought to do in a given situation (Bicchieri and Chavez, 2010).

It is therefore likely that a social gender norm that expects women to be responsible for unpaid work also only applies under certain conditions. Some of the research on gender norms implicitly uses this conditional interpretation and admits that the effect of gender norms is dependent on couples' social situations. For example, Szinovacz (2000, p. 80) states that there is an unwritten rule that expects men to participate in housework "if such help is needed." Situational changes may also alter the impact of gender norms on behavior; for example, marriage and the birth and/or the age of a child are expected to amplify the effect of gender norms on behavior (e.g., Carlson and Lynch, 2013; Grunow et al., 2012; Gupta, 1999; Khoudja and Fleischmann, 2018; Kühhirt, 2012; Sanchez and Thomson, 1997; Schober, 2013; Schulz and Blossfeld, 2006).

In sum, this conditional interpretation of gender norms (albeit often only implicitly argued) differs greatly from the historical, unconditional interpretation of Parsons. The conditionality of gender norms, however, implies that the effect of gender norms is altered depending on the social situation (e.g., the existence of a child or men's and women's earnings potential). Whether gender norms are applied conditionally or unconditionally, they are used to explain the division of labor between men and women. In the next section, I will summarize the theoretical approaches based on gender norms and examine the extent to which the underlying mechanisms are suitable to explaining the relationship between gender and share of housework.

1.3.2 Disentangling the Mechanism Behind Gendered Division of Paid and Unpaid Work

For decades, researchers have attempted to explain the gendered division of paid and unpaid work that has been the status quo in nearly all OECD-countries (OECD, 2017). In essence, two different theoretical arguments can be distinguished that either independently or in combination explain why women are primarily responsible for domestic work and child care, while men are the main earners. The first argument sees gender norms as a driving force for individuals' shares of paid and unpaid work, the second

[10] They describe a medical doctor who might feel obligated to not tell a patient the truth about a cancer diagnosis if convinced that the patient would commit suicide (Diefenbach and Opp, 2007, p. 489; Hechter and Opp, 2001, p. 405f.).

argument focuses on economic factors such as time spent on work or money as explanatory factors for couples' division of paid and unpaid labor.

Gender Norms and Behavior

First, the gender norm approach argues that gender-specific norms directly drive individuals' behavior (see, e.g., Behr et al., 2013; Bittman et al., 2003; Brines, 1993, 1994; Carlson and Lynch, 2013; Davis and Greenstein, 2009). Individuals internalize these prescriptive norms[11] during socialization, which later define individuals' level of support for a division of paid and unpaid work that is based on gender (Davis and Greenstein, 2009). It is expected that individuals who hold traditional gender role attitudes divide paid and unpaid labor traditionally (i.e., a male breadwinner and a female homemaker), while egalitarian individuals opt for a work division that is not based on gender (see, e.g., Bittman et al., 2003; Davis and Greenstein, 2009).

At first glance, this seems to be a reasonable argument, which is likely why it is almost always applied in other research (for overviews, see Coltrane, 2000; Grunow, 2019; Lachance-Grzela and Bouchard, 2010). What remains unclear, however, is why couples act according to internalized gender norms. Or put differently: What exactly is the mechanism through which gender norms affect behavior?

Research has shown a gap between attitudes on behavior in various areas of life (for a recent overview, see, e.g., Thiel, 2020). This phenomenon has been predominantly discussed in research on the effect of environmental attitudes on behavior and has led to the formulation of the low-cost hypothesis (Diekmann and Preisendörfer, 1998, 2003). The low-cost hypothesis argues that the effect of attitudes on behavior varies with the behavioral costs; only if costs are low, an effect on behavior can be expected (Diekmann and Preisendörfer, 1998, 2003). Hence, the mere existence of a social norm is no sufficient explanation for a particular behavior (see, e.g., Rauhut and Krumpal, 2008). Instead, it must be considered that behavioral costs likely moderate the effect of attitude or norms on behavior (see, e.g., Best and Kroneberg, 2012; Braun and Franzen, 1995; Rauhut and Krumpal, 2008). Transferred to the effect of gender norms, an effect of attitudes on behavior can only be expected in situations where costs are low. In high-cost situations, a gender norm concerning the division of paid and unpaid labor cannot be expected to prevail—at least according to low-cost hypothesis. At the very least, it is questionable whether having fewer hours of gainful employment, less work experience, and lower (potential) earnings is really a low-cost situation for women.

In some related areas, this low-cost hypothesis has already been established as an explanation for the gap between attitudes and behavior (see, e.g., Auspurg et al., 2014, who study job-related migration decisions within partnerships). To my knowledge, in

[11] A prescriptive gender norm states what a person ought or ought not to do based on their gender (Jasso and Opp, 1997, p. 948).

the context of gender norms and the division of paid and unpaid labor, costs for gender norm conforming behavior have not been discussed so far. Consequently, the mechanism behind why traditional gender role attitudes affect women's share of housework is not yet comprehensibly explained by existing literature.

Second, the gender deviance neutralization approach argues that individuals need to compensate gender norm violations (Brines, 1993, 1994). At least implicitly, this approach considers certain costs that arise from behavior that deviates from the gender norm. To meet these costs individuals may use, for example, the symbolic potential of housework (Brines, 1994). It is therefore expected that women who deviate from their normative role as homemakers (e.g., by earning more than their male partners) compensate for this norm deviation by taking on most of the domestic work. Similarly, men who deviate from their breadwinner role are expected to decrease their participation in domestic tasks.

What is not discussed, however, is that individuals do not only need to consider potential costs from gender norm deviant behavior; they also need to consider possible gains from this behavior (see, e.g., Akerlof and Kranton, 2000). Instead, Bittman et al. (2003) specifically argue that complying with a gender norm is not caused by "carrots and sticks" (p. 191). They argue that it is a combination of an individual's need to make sense of the world around them and their expectation of others' behavior that causes them to act according to gender norms (Bittman et al., 2003, p. 191). From a rational-choice perspective, this is a rather vague explanation for the effect of norms on behavior.

In contrast, Akerlof and Kranton (2000) apply the rational-choice approach. They assume that individuals follow internalized, prescriptive gender norms (or "prescriptions", as the authors call them) to affirm their self-image or identity as a man or as a woman; violating these norms would lead to anxiety and discomfort.[12] Their argument additionally consists of the assumption that failure to conform to a prevailing gender norms causes social sanctions (e.g., social exclusion), which is in line with general literature on the effects of social norms: "the severity and probability of external sanctions, including informal sanctions, are costly as well" (Opp, 2001, p. 10718). As a result, norm deviation reduces the utility of the norm deviator.[13] As an example, Akerlof and Kranton (2000, p. 747) describe a man who does not provide most of a heterosexual couple's household income, thus deviates from the normative expectation. He thus experiences utility losses, as this deviation not only undermines the (internalized) image of how a man should behave, but can also result in social sanctions. His female

[12] This is in line with Opp (2001, p. 10718), who states that "if norms are internalized this means that breaking a norm ensues in a bad conscience which is a cost."

[13] In more detail, Akerlof and Kranton (2000) argue that individuals experience payoffs not only from their own actions, but also from the actions of others. These payoffs are not considered stable, but can persistently be changed by others.

partner's utility would also be reduced, as she is seen as taking part in the male sphere and neglecting her ascribed "female" tasks. To increase utility, the couple could decide that she (and not he) does most of the housework to compensate for both of their norm deviations.

Moreover, a utility function does not only consist of costs, but benefits must also be taken into account, for example, because certain preferences can be fulfilled (e.g., Akerlof and Kranton, 2000). Women might benefit from an egalitarian work division insofar as they benefit from a high personal income (e.g., in case of a possible separation) and/or from not having to do most of the housework. Also a couple's joint utility might increase if, for example, a relatively better educated woman spends more time in paid work than her relatively less educated partner. Consequently, individuals or couples may only increase their utility by a traditional division of paid and unpaid work if the potential costs to them of deviating from the gender norm exceed the potential benefits of that deviation. Or put differently, if potential benefits from norm deviation are greater than potential costs, individuals can be expected to deviate from the gender norm.

At least when measured with the standard item questions, 70 to 80% of Germans adhere to egalitarian gender norms (Eyerund and Orth, 2019). Therefore, it is questionable whether individuals actually experience costs like punishments or sanctions when deviating from traditional gender norms and, for example, divide housework and child care equally.[14] If traditional gender norms are on the decline, women's costs for deviating from their (former) role as homemakers should also be on the decline.[15] Overall, this would increase the utility of an egalitarian division of work for women. Men, on the other hand, do not benefit as much from a decline in traditional gender norms as their utility not only increases from not deviating against traditional gender norms, but they also benefit from not having to participate in domestic labor. When traditional gender norms are on the decline, one might, therefore, expect women to adapt an egalitarian gender role attitude faster than men as they benefit more from the adjustment (see Bolzendahl and Myers, 2004; Myers and Booth, 2002).

Other Explanations

Economic theories see money or available time as the main mechanisms for explaining the gendered division of paid and unpaid work. These approaches are gender-neutral, meaning that their arguments can be applied for both men and women and do not a

[14] It is conceivable, however, that such sanctions are still possible in some very traditional subgroups, for example those known to have above average traditional gender role attitudes (e.g., lower educated, older individuals; see Mays, 2012). In other, (especially) egalitarian subgroups, one would perhaps even expect sanctions for an unequal division of labor (e.g., pressure to justify oneself in front of others).

[15] For the general argument on social norms, see Opp (2001, p. 10719) who states that "changing norms may change the costs or benefits of various types of actions."

priori see gender as the main driver of couples' division of work.

Approaches such as resource or bargaining theory (Blood and Wolfe, 1960; Ott, 1992) or family economics (Becker, 1965, 1985, 1993; Mincer, 1962) essentially define the acting mechanism as different income potentials of men and women. While resource and bargaining theory focus on individual self-interests (i.e., maximizing own utility), family economics focuses on couples' joint utility. However, both approaches stipulate that the (individual or joint) utility can be maximized if the division of paid and unpaid work is based on the income potentials of the partners so that the partner with the higher income potential is responsible for paid employment. On average, men not only have higher hourly wages than women, but there is also a gender difference in education, especially in older couples (see, e.g., Buchmann et al., 2008; Katrňák and Manea, 2020) as well as an age difference (see, e.g., Gustafson and Fransson, 2015) that allows men to accumulate more work experience. Among others, these differences lead to men having a higher income potential than women.[16] If one partner then decreases his/her hours in paid employment (e.g., to take care of housework and/or children), this is more likely to be the female partner, which would then lead to a traditional division of paid and unpaid labor. Analogous approaches based on time availability (e.g., Presser, 1994) argue that the partner who has more time available (i.e., who works less hours in paid work) takes over the domestic work. As a result, housework and child care are also attributed to women, as men (with their higher earning potential) spend more hours in paid work.

Drawing from theories of distributive justice (see, e.g., Deutsch, 1975), the normative equivalent of letting money and/or available time decide are norms that are based on the justice principle of equity. In short, equity implies that both partners receive outcomes proportional to their inputs (see Deutsch, 1975). In the context of the division of paid and unpaid work, inputs are economic factors such as income or wealth. Providing more of an input allows a partner to reap more rewards in the relationship, for example not having to do domestic work and more time for leisure instead. Hence, the norm of equity expects the partner who earns less or who works less hours in paid work to do more housework.[17]

As with gender norms, the assumption that equity norms affect couples' behavior also requires a mechanism-based explanation. If equity is the prevailing norm in a society, one explanation can follow the argument of Akerlof and Kranton (2000). Assuming that individuals' aim to maximize utility, an equity norm-conforming behavior can be

[16] It is also argued that the fact that women's productivity in care work is higher than men's explains why family economics expects women to focus on child care and housework, even if they have the same earning potential as their male partner (see, e.g., Becker, 1993; Brines, 1994).

[17] Another often-mentioned social norm is the equality norm. A division of work that is based on equality would expect both partners to do exactly the same shares of each task, for example the same hours of housework independent of the situation (i.e., monetary inputs).

expected if the gains from conforming are greater than possible losses from deviation. For example, the partner who earns less might experience losses from having to do the larger part of the housework. However, if gains from an equity norm-conforming behavior are greater than these losses, utility is maximized.

In contrast to the theoretical discussion on whether money or gender drives behavior, most applied research does not argue with such strict differentiations between the two possible explanations for the gendered division of work. Instead, most researchers expect both variables to interact with each other (e.g., Khoudja and Fleischmann, 2018; Killewald and Gough, 2010; Sanchez and Thomson, 1997, and many more). Likewise, it is possible that on the normative level individuals base their expectation for who should be responsible for housework on both gender norms and equity norms. For example, individuals might have internalized the norm that the more money a partner earns (i.e., relative or absolute income), the less time he/she should spend on housework. At the same time, they might also have internalized that there should be a gendered level difference here, so that men still have to do slightly less housework than women with the same relative income. They might also expect men and women to share their duties in the labor market, while they still expect women to be primarily responsible for child care (see Grunow et al., 2018).

In sum, normative explanations based on an unconditional social norm that drives the gendered division of labor do not provide clear mechanisms to explain the observed relationship. Adding explanations based on a rational-choice understanding of human behavior might help to unravel the explanatory mechanisms. To do so, however, the interpretation of social gender norms must move from an unconditional to a conditional interpretation, according to which different gender norms are applied depending on the social situation. In the next section, I argue why the standard measurement of gender norms with item questions is not able to measure gender norms that are applied conditionally, and indicate why this might be problematic for research on the gendered division of labor.

1.4 Reflections on the Standard Measurement of Gender Norms

1.4.1 Criticism on the Standard Measurement

The measurement of gender role attitudes, or rather internalized (prescriptive) gender norms, has been criticized in the past (for a recent overview, see Walter, 2018b). Many of the item questions included in large population surveys are based on scales developed in the 1960s and 1970s that typically represent a traditional gender norm phrased around a role of women that is increasingly less common in Western societies today (Braun, 1998). These original scales can now hardly discriminate between respondents' attitudes as most individuals disagree with the traditionally phrased item questions,

which leads to ceiling effects (Barth and Trübner, 2018; Halimi et al., 2018; Walter, 2018b). As the traditional wording of the item questions does not allow an egalitarian perspective to be expressed, egalitarian respondents might perceive them as outdated or ill-defined (Braun, 2008; Walter, 2018b). As the content of the item questions revolves around the (traditional) roles of women, capturing newer nuances connected to the (changing) roles of men (e.g., men sharing responsibilities for household and/or family work) is hardly possible (Braun et al., 1994; Halimi et al., 2018; Walter, 2018b).

To tackle these criticisms, newer item questions with an egalitarian perspective have been introduced. Examples from these egalitarian item questions include "If a woman can make a better career than her husband, he should help her with her career rather than try to get ahead in a career of his own" or "It is good if the man stays at home and cares for the children and the woman goes out to work" (see Braun, 2008). But also here, there are a number of problems: The egalitarian questions have been criticized for not distinguishing well between traditional and non-traditional respondents (Braun, 2008). Braun (2008, p. 650) argues that the (uncommon) egalitarian arrangements have been less discussed amongst traditional respondents which leads them to concentrate only on single components of the statements, ignore the rest, and, agree with egalitarian-worded items (Braun, 2008). Also, egalitarian respondents do not always agree with the egalitarian items as there are different dimensions of egalitarian gender norms: some prefer equality between genders, others might not want to restrict anyone to housework, while others might prefer individualized solutions (Braun, 2008). This is supported by recent research. For example Knight and Brinton (2017) and Grunow et al. (2018) argue that gender norms no longer appear on a unidimensional scale ranging from "traditional" to "egalitarian"; instead, gender norms should be understood as a multidimensional construct. They assume that previously overlooked dimensions are, for example, beliefs about intensive mothering (i.e., supporting that women are part of the labor market but simultaneously ascribing women to be care takers of children) or egalitarian essentialism (i.e., preferring an individual solution for each couple). It is unclear, how suitable the global standard item questions are to disentangling these different dimensions of gender role attitudes.

Research has found that the wording of classical item questions does not always allow an unambiguous interpretation of respondents' ratings (Braun et al., 1994) as different individual aspects of each statements could be confounded. Braun et al. (1994) explain this with the common item question "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work." On the one hand, the item may be interpreted as being concerned with the emotional needs of a child if the mother is absent, on the other hand it can be associated with the capabilities of a working mother (Braun, 1998; Braun et al., 1994). Detangling those different interpretations is not possible with just the one item question. Therefore, the

resulting measurement cannot clearly state what agreeing or disagreeing with the item actually means.

In addition, differences in experiences and/or living conditions between countries or regions might also lead to different interpretations of the item statements (Braun et al., 1994). For example, whether the child suffers from their mother working might be influenced by the actual availability of alternative child care and/or by own experiences with working mothers (Braun et al., 1994). In Eastern Germany, more external child care is available compared to Western Germany, while it is also much more common that mothers work. Therefore, respondents in Eastern and Western Germany probably also differ in their interpretations of the item questions. This makes comparisons of gender role attitudes between these two regions difficult (Braun and Scott, 2009). Van Vlimmeren et al. (2017) found that meanings of items regarding family values and gender roles differed across countries, so that they do not represent the same concept across cultures with cluster analyses of 47 European countries of the European Value Survey. Braun and Scott (2009) reported similar results when comparing Anglo-Saxon and former socialist countries.

Despite widespread criticism, standardized item questions remain the status quo for measuring internalized gender norms and their effect on the division of paid and unpaid work among couples. To the best of my knowledge, an additional shortcoming of the item questions has not been discussed before: Auspurg and Düval (2022, see chapter 2 of this dissertation) argue that classical item questions do not capture the conditionality of gender norms, which is likely to bias results.

1.4.2 Measuring the Conditionality of Gender Norms

Against the backdrop of the literature on social norms, Auspurg and Düval (2022) argue that most classical measurements of internalized gender norms have overestimated the prevalence of (traditional) gender norms, and, in turn, their influence on behavior. For the following argument, see also Düval and Auspurg (2018, 2019). The main argument here is that standard item measurements do not capture the conditionality of gender norms, as respondents can only agree or disagree with broad, global statements such as "Housework should be equally split between men and women." There is no (further) information about the social situation of the men and women addressed. A measurement like this only works if respondents' answers—and therefore gender norms—are unconditional of the (hypothetical) social situation.

If gender norms are applied conditionally, however, they would only have limited applicability and be bound to the particular social situation in which they are to be applied. Therefore, to decide whether one agrees with the previous statement, information on the couple's social situation is required (see Auspurg and Düval, 2022). For example, there is no information on the employment situation of the described

couple in the item question mentioned. Thus, measuring a conditionally applied gender norm with a standard item question such as this is hardly possible. As an aside: standard item questions also fail to confirm the unconditionality of gender norms; it is simply assumed.

So, what do respondents do when necessary information is missing from the standard item question? One possibility is to skip the question. As item non-response is typically low for standard item questions,[18] most respondents likely opt for another option. Respondents may (likely unconsciously) fill in additional context to the briefly described statement. Auspurg and Düval (2022) suggest that respondents do not impute just any social situation to create a context, but a social situation that is most likely the case. Typically, in Western societies, men and women divide their work traditionally. When faced with a judgment of whether men and women should do the same amount of housework, respondents likely consider a typical couple with a traditional division of work. The measured internalized gender norm is then conditional to this imputed behavior, and therefore more of a descriptive gender norm (*what is most likely the case*) than the intended prescriptive gender norm (*what should be the case*) (Auspurg and Düval, 2022).

In addition, a measurement instrument that ignores the conditionality of gender norms can lead to confusion between gender norms and equity norms. Usually, respondents' gender role attitudes are inferred from their responses to item questions. For example, agreeing with the statement that *women* should be responsible for housework and/or child care is equalized to having traditional gender role attitudes. And also disagreeing that *men* should be responsible for the unpaid work is interpreted as a traditional gender role attitude. In the end, the gender of the person described in the item question is seen as the cause for a change in respondents' response behavior (agreeing vs. disagreeing, see dotted arrow in Figure 1.3).

As mentioned, Auspurg and Düval (2022) argue that it is likely that respondents imagine a stereotypical (traditional) social situation when answering item questions. If so, the gender in the item question leads to stereotypical beliefs about the described individual's behavior: women are assumed to be responsible for domestic work and child care, while men are assumed to be the employed main earners (see solid arrow in Figure 1.3). In other words, the effect of the described person's gender on the response behavior is mediated by respondents' beliefs on how a typical couple divides their paid and unpaid work (i.e., descriptive gender norms). As a result, it is not (necessarily) gender or prescriptive gender norms that drive the evaluation of an appropriate division of work. If assumed that men have higher earnings and more employment hours than women, individuals who adhere to equity norms would also agree that men should do

[18] For example, only 1 of 4,750 respondents did not provide a valid reply to the item question "Men should participate in housework to the same extent as women" (pairfam 2017/18).

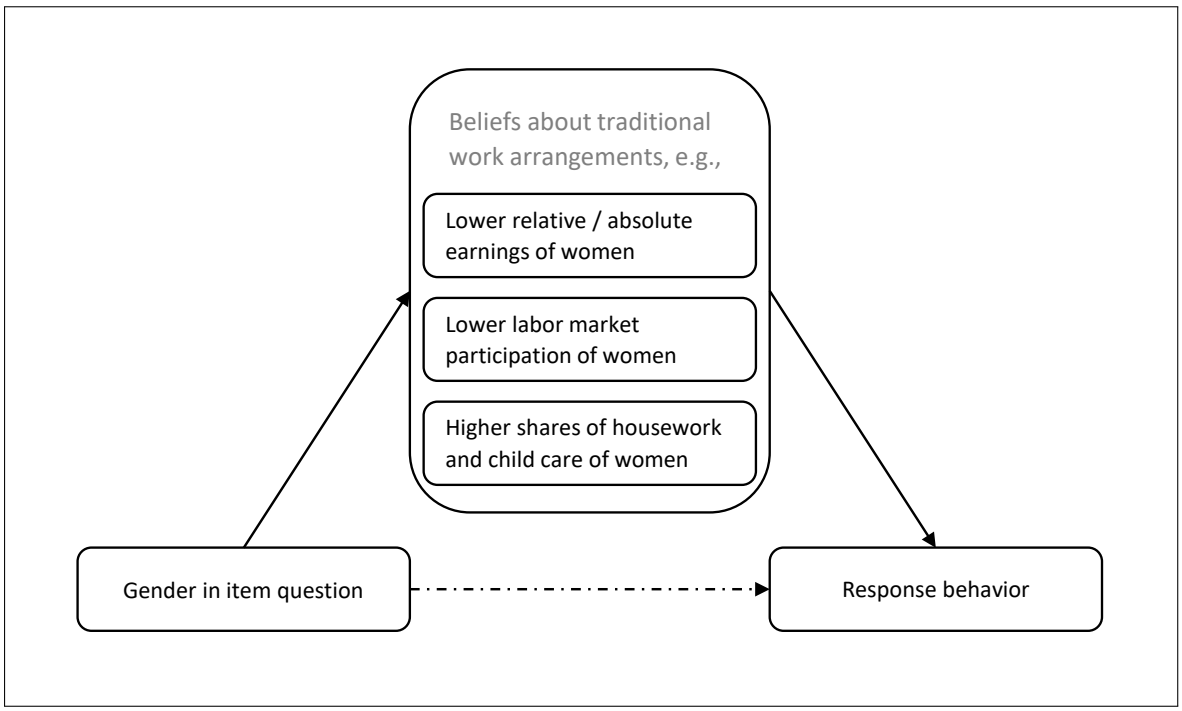


Figure 1.3. Effect of Gender Presented in the Item Question on Response Behavior

Note: This figure shows the assumed causality underlying most literature using standard items to measure gender norms. The response behavior is attributed to the gender presented in the item question (dotted arrow). Auspurg and Düval (2022) argue, however, that this correlation could be mediated by respondents' stereotypical beliefs about traditional work arrangements. These beliefs could—together with equity norms—give a strong impetus to agree with traditionally worded item questions, even when respondents do not adhere to gender norms (solid line). This makes it impossible to distinguish whether respondents agree to traditional item questions based on gender norms or equity norms.
Source: Own illustration. See also Auspurg and Düval (2022).

less of the domestic work than women. As a result, when researchers find that gender has an impact on who should do the housework (or paid work), this can no longer be interpreted as support for traditional gender norms if only the standard item questions are used to measure gender norms.

If this argument holds, it is very likely that gender role item questions confuse prescriptive gender norms, descriptive gender norms, and equity norms. In addition, it is not distinguishable if group differences (e.g., between men and women) in measured gender role attitudes are caused by actual differences in prescriptive gender norms, or if sub-groups simply differ in their imputations of what they think is most likely the case (i.e., differences in descriptive gender norms). In other words, if respondents differ in the social situations they imagine in order to answer the item questions on gender role attitudes, this would lead to unobserved heterogeneity between respondents. Not controlling for this unobserved heterogeneity biases the measured gender role attitudes, or rather the measured differences in gender role attitudes.

1.5 The Current Contribution

1.5.1 Research Gap and Objective

In the last sections, I pointed out several issues concerning gender norms as an explanation for couples' (unequal) division of paid and unpaid work. In the following, I identify the research gap that arises from the aforementioned issues and explain how I plan to fill this gap with this dissertation.

How do couples decide how to divide paid and unpaid work? I have argued that two widely used approaches, the gender norm approach and the gender deviance neutralization approach, do not provide comprehensible mechanisms to explain this behavior. Instead, both approaches mainly assume that individuals' attitudes must affect their behavior or that individuals follow gender norms because they act according to what they believe are the others' normative expectations of what a woman/man should do. I have argued that the explanations of both approaches remain rather vague. Akerlof and Kranton (2000) introduced the rational choice perspective to research on gender norms. In this view, the division of paid and unpaid labor among couples is only driven by a social norm if non-compliance with this norm entails social sanctions.

In addition, the conditionality of social gender norms is substantial, not only for the theoretical understanding of the mechanism behind gender norms, but also for appropriate measurement approaches. I have argued that existing measurement methods relying on item questions only allows "unconditional" gender norms to be measured. However, as respondents' judgments do not only depend on internalized (gender) norms, but also on the social situation imputed by respondents when responding to item questions, it is not possible to disentangle whether respondents apply gender norms or gender-neutral norms to rate the broad statements used to measure gender role attitudes.

In sum, this dissertation aims to uncover whether gender norms can (still) be considered a relevant mechanism for explaining the relationship between gender and the division of paid and unpaid work. To this end, I have introduced the concept of the conditionality of social gender norms, which is important in both theoretical and methodological arguments. The goal of this research is to detect if and under which conditions it is believed that women should be held responsible for housework and child care.

In more detail, I probe the (un-)conditionality of gender norms: Is there a social norm that expects a gendered division of paid and unpaid work independent of the social situation? Or does a conditional gender norm prevail, which expects couples to (also) divide their paid and unpaid work based on their social situation (e.g., their employment characteristics)? Are there more traditional sub-groups that apply gender norms unconditionally, while other sub-groups might apply them conditionally (e.g.,

men vs. women)?

Assuming that gender norms are applied conditionally, I also investigate the social situations in which women are expected to be responsible for the majority (if not all) of housework. Are the trade-offs of paid work and unpaid work the same for men and women (i.e., equity)? Or is there the normative expectation that employed women should do more housework than men in the same employment situation? Is there even a social norm that expects women to do the majority of the housework in situations in which they deviate from their normative role as homemakers (e.g., when earning relatively more than their male partners)? And again, are there sub-groups that differ in their normative expectations of which condition women should be responsible for doing housework?

1.5.2 Analytical Approach

Typically, agreement with gender role item questions is equated with agreement that members of a particular gender should behave as described in the item. From this, the individuals' gender role attitudes or their internalized gender norm are inferred. In other words, a genuine effect of gender on item agreement is hypothesized that is neither confounded nor mediated by other variables.

The main critique that I introduced in this dissertation is that this approach ignores that gender norms are applied dependent of context. My argument is that individuals who should judge whether a described behavior conforms to their normative expectations need extensive information on the social situation the behavior occurs in. If the necessary information on the social context is not provided, respondents impute it based on a relevant predictor. As gender is correlated to, for example, employment situation, it is likely that respondents use the gender of the female and male partners described in the items to impute their employment hours and their (relative) income.[19] The described behavior is then not only judged based on gender, but also on the imputed employment situation. In other words, the imputation of employment hours mediates the effect of gender on the agreement with the item question, which contradicts the assumed causal effect of gender in the item questions.

Not considering this mediation only allows the effect of gender on agreement with the item question that is fixed to the imputed social situation of the described couple (e.g., employment hours or relative income) to be measured. As a result, it is not possible to disentangle whether women should do more housework a) because of their female gender (direct effect) or b) because based on gender, respondents impute lower employment

[19] In Germany, for example, the typical couple is comprised of a male partner who works full-time and a female partner who works part-time (e.g., German Federal Statistical Office, 2022b; Gartner and Hinz, 2009; Nitsche and Grunow, 2016). Typically, he earns about 20% more per hour than she does, while she is mainly responsible for housework and child care (e.g., German Federal Statistical Office, 2022b; Nitsche and Grunow, 2016).

hours or income for women (indirect effect). The former suggests the presence of traditional gender norms, whereas the latter indicates other social norms such as equity norms. Consequently, it is not possible to conclude that an individual has a traditional attitude just because he/she agrees with traditionally worded item questions on gender norms. Instead, causal mediation analysis is necessary to disentangle the two paths.

This dissertation builds on the suggestion of Acharya et al. (2018) to use a factorial design that can disentangle the total and direct effects of gender by using a two-step experimental approach. In the first step, the treatment of interest (i.e., gender in a hypothetical statement) is randomly assigned. In this way, I can measure the average treatment effect. Imagine a gender role item question where instead of "man" or "woman" there is a placeholder which randomly switches between the two genders. Information on possible mediators (i.e., employment situation) is not provided. The mediators therefore take their natural values, i.e., the values respondents impute based on the gender randomly assigned to the placeholder. For example, if the placeholder changed to "man," respondents would assume full-time employment, while "woman" would lead respondents to assume part-time employment.

In the second step, information on possible mediators (i.e., employment situation) is provided. This information varies randomly in its content (e.g., the female partner works either 20, 30, or 40 hours per week). Respondents no longer need to impute the values of the mediators, but can base their judgments on the existing information. This allows the controlled direct effect of the treatment (i.e., the part of the total effect of gender that is due to neither mediation nor interaction with the mediator) to be measured. In other words, the remaining effect of gender can be measured when the mediator (e.g., employment situation, other relevant mediator) is fixed to a given value. Picture a situation where the male and the female partner both work 40 hours per week and earn approximately the same. If an effect of gender on the appropriate housework share remains, this would be indication for a traditional gender norm.

In addition, the existence of equity norms can be probed. Imagine that an effect of gender was found to be unequal to zero in step one. Now, this effect disappears in step two after providing the extensive contextual information (e.g, the information that both partners earn the same). The total effect of gender that was found in step one is then completely mediated by the couple's financial situation. If so, one could conclude that respondents do not base their judgments on gender, but instead consider the employment situation: given equal earnings, they feel that both partners should do the same amount of housework, which would support the existence of equity norms.

I use a factorial survey experiment that follows the suggestion of Acharya et al. (2018) by design. The factorial survey experiment relevant for this dissertation was included in the 10th wave of the German Family Panel pairfam (Brüderl et al., 2019). Düval and Auspurg (2020) designed the experiment and ran several pre-tests before it

was included in the pairfam questionnaire.[20] For more information on the experimental set-up, the sample, first descriptive results, and data quality, see Düval and Auspurg (2020).

The factorial survey experiment allows for an investigation into whether and why respondents believe that women should do the (majority of) housework. First, I examine on the (un-)conditionality of gender norms. Second, I dig deeper into the social situations in which women are (not) expected to do housework. To do so, I also consider that the results may differ between more or less traditionally oriented sub-groups (e.g., men and women).

1.5.3 Methodological Considerations

Instead of standard item-based measurements, I use a factorial survey experiment to provide respondents with extensive contextual information on a couple's work-family situation in order to adequately measure gender norms. Although I argue that factorial survey experiments are better-suited to measuring gender norms than the standard item-based measurements, unintended effects of the chosen survey mode could also bias the results of factorial survey experiments. Answering a factorial survey module is cognitively more challenging than answering item questions. Learning and/or fatigue effects may play a role insofar that repeated evaluation of similar vignettes could influence response behavior. For an overview of these and other possible method effects, see, e.g., Auspurg and Hinz (2015) and Sauer et al. (2011).

Research has shown that to avoid possible method effects and to account for respondents' cognitive limitations, no more than 10 vignettes should be posed (Auspurg and Hinz, 2015). Similarly, vignettes should contain no more than 7 (+/- 2) dimensions (Auspurg and Hinz, 2015). The factorial survey experiment relevant to this dissertation consisted of 3 vignettes per respondent with a maximum of 7 variable dimensions in each vignette.

Figure 1.4 shows an additional method effect that could bias the main results presented here. I have argued that the main advantage of a factorial survey experiment is that it allows for the disentanglement of different factors that influence respondents' evaluations of an appropriate share of men's and women's housework or total workload (see Figure 1.4, grey arrows). Not only is the treatment variable (i.e., gender) varied randomly, but also possible mediator variables (i.e., information on couples' work arrangements). With this approach, I can identify a direct effect of gender if the mediators are fixed, or rather controlled, at certain levels of the vignette dimensions.

[20] The authors included an additional experiment into the pairfam survey. Here, not only the pairfam respondents, but also their partners answered the same factorial survey experiment. This experiment is not part of this dissertation. Information on the experimental set-up can be found in Düval and Auspurg (2022).

However, this identification of the direct effect is only possible if results are not biased by method effects. However, it is possible that design features of the factorial survey experiment affect results. For example, the order in which dimensions are presented may moderate the effects on respondents' evaluations (see Figure 1.4, black arrow).

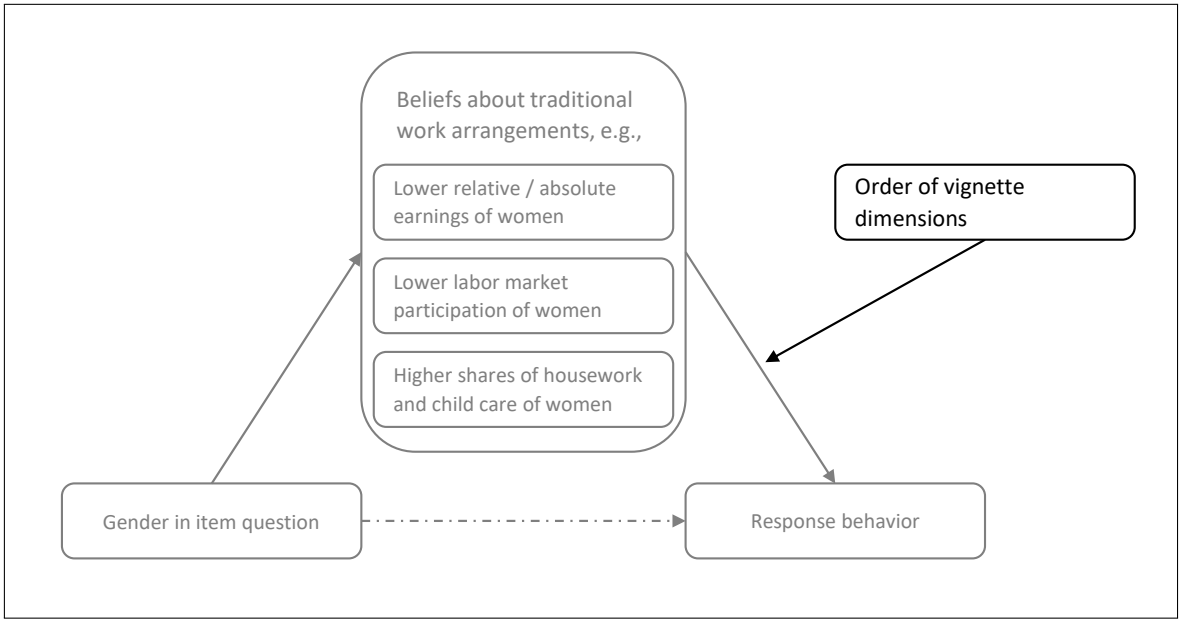


Figure 1.4. Effect of Order of Vignette Dimensions

Note: This figure shows the assumed interaction (i.e. moderation) between the order of the vignette dimensions and the information on the couples' work arrangements on the response behavior (black arrow).
Source: Own illustration.

Respondents must weigh a certain number of dimensions in order to evaluate a factorial survey experiment. The order in which vignette dimension are presented may have an effect on the response if respondents do not consider all dimensions. Especially for older or less educated respondents, it can be a cognitive challenge to capture all dimensions. Likewise, very fast respondents might not take enough time to process all dimensions as they might stop reading after finding a plausible justification for their answer, or only remember the last dimensions they have read. In either case, this would result in primacy (i.e., only the first few dimensions are considered) or recency (i.e., only the last few dimensions are considered) effects.

Auspurg and Jäckle (2017) were the first to show that effects of dimension order can occur in factorial survey experiments, especially when the task is complex. Robbins and Kiser (2018), on the other hand, did not find any evidence for dimension order effects although both studies used data from homogenous samples of university students. As it is likely that order effects are stronger for less educated and/or older respondents (e.g., Sauer et al., 2011), it is unclear whether the factorial survey experiment relevant for this dissertation might still be biased by dimension order effects. The pairfam survey

is a general population survey of three age cohorts with a relatively young sample—the oldest respondent of the factorial survey experiment was 47 years old in wave 10. However, different levels of education could play a role here.

Therefore, to ensure that the main results of this dissertation are not biased by method effects, I carefully consider one limitation, namely whether the order in which the information is presented in the factorial survey experiment affects the results of the experiment.[21] For this reason, I use data from a general population survey to examine whether the order of the vignette dimensions used has influenced respondents' ratings.

1.6 Key Findings

1.6.1 My Contribution to the State of Research

The findings in this dissertation are based on three different research articles (see Table 1.2), each with its own research focus. As a whole, they are intended to help close the research gap underlying this dissertation. The first article (chapter 2, Auspurg and Düval, 2022) was written in collaboration with Katrin Auspurg. A factorial survey experiment was created and implemented as an experimental module into the German Family Panel pairfam. The factorial survey experiment itself was created in close collaboration with my co-author. In addition, I conducted pretests, discussed the module with the pairfam study's principal investigators, and coordinated the implementation of the module into the pairfam survey. Although data analysis was conducted by both myself and my co-author, the final results are based on my calculations. Katrin Auspurg took the lead in writing the drafts of the introduction, theory, and state of the research, whereas the remaining parts of the article (especially the sections on data and methods, results, discussion, and the abstract) are mainly my work. We both revised each individual section, but the final review and revision of the manuscript was my task.

The factorial survey experiment was also used to answer the research questions from the second article (chapter 3, Düval, 2023). All elements of this manuscript are exclusively mine: Guiding idea, data analysis, writing, and final review.

The third article (chapter 4, Düval and Hinz, 2020) is based on an idea developed with the support of Katrin Auspurg. The research questions were answered based on a factorial survey experiment that was prepared and implemented by my co-author Thomas Hinz. I conducted the data analysis independently. The introduction, state of the art, theory, data and methods, and results sections were written by me. Thomas Hinz wrote the abstract and the discussion, and conducted the final review of the

[21] Of course, dimension order is only one possible method effect that could bias the main results of this dissertation. The presentation format (i.e., running text "vignettes" vs. tabular "profile cards"), the number of dimension levels, or even an interaction of all these three design features might also play a role here (for this argument, see, e.g., Thiel et al., 2021).

Table 1.2. Overlook of Articles and Own Contribution

Ch.	Bibliography	Own Contribution	Weight	Score
2	Auspurg, Katrin & Sabine Düval (2022): Probing on the Conditionality of Gender Norms: Is the Significance of Gender Overrated? *Unpublished Working Paper.*	65%	1	65
3	Düval, Sabine (2023): Do Men and Women Really Have Different Gender Role Attitudes? Experimental Insight on Gender-Specific Attitudes toward Paid and Unpaid Work in Germany. *Social Science Research*[a]. 112:1-20.	100%	2	200
4	Düval, Sabine & Thomas Hinz (2020): Different Order, Different Results? The Effects of Dimension Order in Factorial Survey Experiments. *Field Methods*[b]. 32(1):23-37.	80%	1.5	120
				385

Note: [a] 2-year Impact Factor: 2.617; 5-year Impact Factor: 3.416 (Journal Citation Reports 2021).
[b] 2-year Impact Factor: 1.782; 5-year Impact Factor: 3.156 (Journal Citation Reports 2021).

manuscript. I was responsible for the revision of the article after receiving reviews from the journal it was published in. The response to the reviewers was written by Thomas Hinz based on my notes.

In the next section, I will use the main findings from the three research articles to fill the research gap introduced in this first chapter of my dissertation. In doing so, I will not summarize each article on its own, but rather discuss the key findings based on the three research aims I posed in section 1.5:

1) Are gender norms applied (un-)conditionally?

2) Under which social situations are women (not) expected to do the housework?

3) Does the order of the vignette dimensions influence respondents' ratings?

1.6.2 Why Should Women Take Care of the Housework?

Is there (still) a gender-specific norm that expects women to do housework and child care? Based on the interpretation of the factorial survey experiments used in chapters 2 and 3 of this dissertation, the answer is quite clear: No, there is not. By only using standard item questions to measure gender norms, however, the result would not have been as clear. With about 20%, a relevant number of respondents does not agree with the statement that "Men and women should participate in housework to the same

extent as women". Still, with the mean being 4.3 (standard deviation [s.d.]: 0.9) on a 5-point rating scale that ranges from 1 "Disagree completely" to 5 "Agree completely", most respondents actually agree with the statement.[22] I argue, however, that it is not clear as to why respondents agree or disagree with this statement. Do respondents actually base their evaluations on the genders presented in the item question? At least the existing literature interprets the results in this way. Respondents who agree that men and women should share housework equally are assigned an egalitarian attitude, respondents who disagree with the statement are assigned a traditional attitude.

Chapter 2 of this dissertation (Auspurg and Düval, 2022) challenges this interpretation of respondents' response behavior. By using a factorial survey experiment, we were able to disentangle the two explanatory paths of respondents' response behavior: 1) the direct path from the gender presented in the item question to respondents' response behavior, and, 2) the indirect path mediated through respondents' beliefs about men's and women's typical work arrangements. The general approach has been described in section 1.5.2; in short, only the existence of a direct effect of gender after controlling for relevant mediators (i.e., the described couple's employment situation) would point toward the existence of gender-specific norms. In a first step, we used an experimental design in which only the gender of the person described is varied randomly. Under this condition of no additional information, respondents classified as "traditional" based on the standard item questions rate the factorial survey experiment in a way that shows they expect women to be responsible for a higher share of housework than their male partners.[23] As expected, "egalitarian" respondents do not want men's and women's shares of housework to differ. This replicates the findings of the standard item question on gender norms, in which both respondents groups differ in their response behavior.

In a second step, information on relevant mediator variables was provided to respondents, including information on the amount of labor market hours and income of both partners. In this variation, "traditional" respondents' response behavior changes: The effect of the vignette person's gender on response behavior is substantially reduced and no longer statistically significant. Further results suggest that when no information on the couples' labor market characteristics is provided (i.e., the standard way to measure gender norms), "traditional" respondents assume a rather traditional work division between male and female partners.[24] However, if they are provided with information

[22] This item is part of the German Family Panel pairfam (2017/18). For detailed results, see chapter 2.

[23] To classify respondents into the three subgroups "egalitarian," "traditional," or "neither nor," Auspurg and Düval (2022) use a standard item questions on how men and women should divide their housework.

[24] More specifically, the experimental design allows to estimate for which level of the labor market characteristics the gender effect is on par with the gender effect of the "no information" condition. In other words, this approach allows to estimate the natural values of the mediators (i.e., relative

on the couples' labor market characteristics, they condition their evaluations on this information and not on their stereotypical assumptions of a typical couple. In contrast, "egalitarian" respondents have already assumed similar employment situations for men and women; providing the information that, in fact, the couple does (on average) have the same labor market hours on income does not change their evaluation much.

Similar results can be found in chapter 3 of this dissertation (Düval, 2023). I compared men and women as two subgroups known to differ in their internalized gender norms, at least according to previous research based on the standard item questions on gender norms. Instead of the appropriateness of a fictional couple's division of housework, respondents now were asked to evaluate the appropriateness the division of total workload, which includes not only housework but also paid work and child care. Again, I found that if men and women are provided with extensive information on a couple's work arrangement, there is no difference between the two respondent groups. The effect of the described vignette person's gender on men's and women's response behavior is also very small, and there is no substantial indication that men or women unconditionally apply gender norms when evaluating couples' division of work.

As a first interim conclusion, there is no evidence for the existence of an unconditional gender norm. Even among subgroups classified as "traditional" by the item questions or known to be traditional (i.e., men), their evaluation of who should be responsible for housework is conditional on information concerning couples' labor market characteristics. The question however remains as to under which conditions women are held responsible for the (majority of the) unpaid work.

Chapter 2 takes a closer look into the conditions under which a woman's share of housework is believed to be appropriate. Equity norms appear to be the main factor influencing whether women's shares of domestic work is considered too low, too high, or appropriate. This result does not only prevail for the sub-group of "egalitarian" respondents, but "traditional" respondents also apply equity norms when judging the appropriateness of a couple's division of housework when extensive labor market information is provided. More paid working hours and higher relative earnings allows a partner to reduce their time spent doing housework, at least according to respondents' normative expectations of a couple's appropriate division of work. Remarkably, this is the case for male and female vignette persons, indicating that there is again no evidence that respondents expect women to do more of the housework then men under the same conditions.

Chapter 3 provides more evidence for the lack of a gender-specific norm concerning domestic tasks. I found the same pattern when looking at respondent gender differences in regard to the appropriateness of a (hypothetical) couple's division of their total

income and labor market hours), and thus respondents' beliefs of the typical shares of paid and unpaid work performed by men and women. For more information, see chapter 2.

workload (i.e., shares of paid and unpaid work). Men and women do not differ in the factors they believe should influence an appropriate division of paid and unpaid work among couples. For both genders, the appropriateness of a partner's share of the total workload depends on his/her relative amount of housework, child care, hours in the labor market, and income. Again, the gender of the vignette person did not play a substantial role. In line with chapter 2, no evidence for a social norm that expects women to do a larger share of the unpaid work in situations in which they deviate from the normative expectation as the homemaker was found.

So, why should women (still) do the housework? According to the main results of this dissertation, women should (only) do the housework if they provide less of the household income and/or if they work less hours on the labor market than their male partners. If women are involved in the labor market to a greater extent than their male partners, the normative expectation is that the male partner, in fact, should then be responsible for the unpaid work. With a measurement tool that takes the conditionality of social norms into account, gender no longer has an influence on response behavior. Instead, respondents apply the norm of equity.

Of course, this key finding is not unchallengeable. Theoretical and also methodological aspects may question the results and their interpretation. Chapters 2 and 3 provide a number of robustness and validity checks that largely confirm the main findings. However, it should be discussed that evaluating a factorial survey experiment is cognitively more challenging than rating item questions (see section 1.5.3 for this argument). To address this, this dissertation focuses on one possible source of bias: the order in which the vignette dimensions are presented to the respondents.

This bias cannot be tested with the factorial survey experiment that underlies the main results of this dissertation as there was no experimental variation of the order of the vignette dimensions. Therefore, Düval and Hinz (2020) use a factorial survey experiment on allocation decisions in education that was included into the Konstanzer Bürgerbefragung[25] in 2011. As the Konstanzer Bürgerbefragung is also a survey on the general population, a possible bias found in this experimental data is also relevant for the interpretation of the results of the pairfam experiment. In addition, the degree of complexity of both experiments is comparable, as both factorial survey experiments included seven dimensions varied within respondents. Also, respondents evaluated a similar number of vignettes in both experiments.[26]

Chapter 4 tests for overall order effects and recency effects (i.e., do respondents focus on the last dimension that presented in a vignette?). It is considered that specific subgroups might be more prone to order effects, e.g., respondents with lower cognitive

[25] The Konstanzer Bürgerbefragung is an annual panel survey of respondents that live in the German city of Konstanz (see Hinz et al., 2012)

[26] In the Konstanzer Bürgerbefragung, respondents evaluated four vignettes; in the pairfam experiments, they answered three.

abilities or those with very fast completion rates. Overall, there is no systematic evidence for order effects in factorial survey experiments of this medium complexity (i.e., vignettes with seven dimensions). There is, however, some small indication that factorial survey experiments might be prone to (small) recency effects, especially so if respondents answered the vignette module quickly.

Although it is prudent to keep in mind that method effects can always cause a bias, which is not only true for the present dissertation but also transferrable to most research using observational and/or experimental data, it is not an alarming result in terms of the main results of this dissertation. Chapter 4 concludes with the suggestion that when constructing a factorial survey experiment, one should avoid having the dimension with the highest supposed influence at the very end of a vignette and that paradata (e.g., response time) should be used to validate findings. Both suggestions have been taken into account in chapters 2 and 3. The last dimension in pairfam factorial survey experiment contains information on the division of child care within the couple, which is undoubtedly relevant, but admittedly not as central as the vignette person's gender, time spent in housework and paid work, and relative earnings. In addition, both chapter 2 and 3 discuss a possible bias that might be caused by very fast respondents in the sections on validity and robustness; however, excluding these possible "satisficers" from the main analyses does not change the conclusions. Most importantly, even if a dimension order effect was present, it is unlikely that respondents who unconditionally favor a traditional division of work would artificially disregard the gender of the vignette person in their evaluation only because another dimension is presented last.

1.7 Synthesis

1.7.1 Review of Theoretical and Methodological Arguments

The main theoretical finding of this dissertation is that gender norms have been overrated in explaining couples' division of paid and unpaid work. At least for respondents aged 23 to 47 living in Germany, I found broad evidence for the existence of a widespread norm that it is not gender, but relative to income and time availability that should determine how paid and unpaid work is divided among heterosexual couples.

In more detail, I questioned the theoretical mechanism behind the gender norm approach which is regularly used to explain the division of paid and unpaid work among couples (see, e.g., Coltrane, 2000; Davis and Greenstein, 2009; Grunow, 2019; Lachance-Grzela and Bouchard, 2010). I have criticized that most research on the division of paid and unpaid work focuses only on testing hypotheses around behaviors that arise from these competing theories, whereas the mechanisms behind the theories are not questioned. For this reason, I focused on the mechanism behind the gender

norm explanations and discussed whether they are (still) fit to explain the gendered division of work.

I argue that based on findings from existing literature, it is unclear why gender norms or gender role attitudes would affect couples' behavior. Only by including the argument that social sanctions may arise in cases of non-compliance, gender norms might work as a mechanism behind couples' division of work (see Akerlof and Kranton, 2000). However, at least in liberal democracies like Germany, in times of (almost) equal participation of men and women in both the education system and the labor market, alongside widespread support for measures aimed at gender equality in society, it is questionable whether one actually experiences strong social sanctions for deviating from a traditional gender norm on division of paid and unpaid work. Of course, in some traditional subgroups, deviation from the gender norm may still lead to social sanctions, but the existence of a widespread gender norm governing the division of paid and unpaid work among couples seems unlikely.

From a methodological point of view, this dissertation has shown that for an accurate and reliable measurement of a social construct such as gender norms, it is essential to cover all theoretical aspects of gender norms. If important theoretical aspects (e.g., the conditionality of social norms) are omitted and thus neglected in the measurement process, this leads to measurement bias.

More specifically, I argue that most research on gender norms excludes the conditionality of social norms, in which a norm is applied differently depending on the social situation (see, e.g., Axelrod, 1986; Elster, 2018). While some studies implicitly consider the conditionality of gender norms in their theoretical argumentation, they do not account for it in their measurement(s). Further, I argue that a factorial survey experiment is able to measure not only the presence of a widespread traditional gender norm in Germany, but also (and in contrast to the standard item questions) the possibility of the conditionality of gender norms. In addition, subgroup differences between traditional and egalitarian respondents or men and women can be measured and studied in more detail. As the measurement of gender norms by way of a factorial survey experiment is not infallible, I also focus on one possible limitation, namely if the order in which the vignette dimensions are presented to the respondents affects the results of the factorial survey experiment.

In conclusion, I have highlighted the importance of thoroughly examining an often-used theoretical concept to explain a (causal) relationship of interest; especially so when that concept is as popular as gender norms in explaining couples' division of work. As shown, the popularity of a theoretical concept does not necessarily imply the existence of a sound mechanism and/or valid measurement of that concept.

1.7.2 Main Results and Policy Recommendations

I found no evidence for the existence of a gender-specific norm concerning the division of work among couples in Germany. Instead, I found that only under the condition that the female partner earns less and/or works less hours than her male partner do respondents expect her to be responsible for the unpaid work. However, if she has the same or even higher labor force participation than her male partner, the normative expectation is that the couple should share their unpaid work or that he should do more. This result cannot only be found for respondents that have been classified as egalitarian using the standard item measurement of gender norms, but also traditionally classified respondents do not use gender as a criteria to judge the appropriateness of a couple's division of paid and unpaid work if they receive extensive information on the couple's work-family situation. Under this condition, well-known respondent gender differences in gender norms are no longer detectable.

Hence, from a more policy-driven perspective, if gender equality is the goal, further enhancing egalitarian gender norms might not be the appropriate path, as most individuals already support an equitable division of paid and unpaid work between men and women. Based on the findings of this dissertation, it is likely that if women increased their time in the labor market relatively to their male partners, this would increase the normative expectation on men to increase their time in unpaid work. Therefore, supporting women to work full-time in the labor market, decreasing the gender pay gap, and supporting men in decreasing their time in the labor market or taking parental leave might be the keys to more gender equality in work division.

1.7.3 Research Desiderata

There are two possible desiderata for future research. On the one hand, there are possible methodological issues with factorial survey experiments that should be considered. On the other hand, there are substantial arguments on which future research could build on.

As mentioned, evaluating a factorial survey experiment is cognitively challenging, which could lead to cognitive overload (e.g, Auspurg and Hinz, 2015). A relevant number of methodological research on method effects in factorial survey experiments has been published (see, e.g., Auspurg and Jäckle, 2017; Sauer et al., 2011, 2020), and I focus on one other possible limitation in this dissertation: the order of vignette dimensions. There is, however, another possible issue that might be relevant to the findings presented here. Previous research has found that including more vignette dimensions into a factorial survey experiment artificially reduces the effect sizes of all dimensions due to cognitive overload (see Auspurg et al., 2009). My main methodological argument was that providing information on a couple's labor market characteristics decreases the

effect of vignette person gender, explained by the fact that respondents do not need to impute stereotypical beliefs about typical work-family arrangements if this information is provided in the vignettes. However, if the argument of Auspurg et al. (2009) is correct, adding additional dimensions (i.e., information on the labor market characteristics) may also artificially reduce the effect sizes of all other dimensions. This would include a reduction of the effect of gender of the vignette person.

It seems unlikely that switching from four to six or seven dimensions would greatly increase the cognitive effort required of respondents.[27] At the very least, seven dimensions are exactly what has been proposed for research with factorial survey experiments (Auspurg and Hinz, 2015). I would even go as far as to say that it is even more unlikely for respondents that have internalized traditional gender norms to pay less attention to gender of the vignette person only due to the inclusion of additional information. Nevertheless, future research should focus more on the possible artificial effect of adding more vignette dimensions and how this could be avoided.

In addition, there are substantial arguments that offer starting points for new research. First, the factorial survey experiment in this dissertation only contains information on the amount of housework a partner does, not on the specific task. Respondents could assume that male vignette persons predominantly do male-assigned tasks (i.e., repairs and maintenance), whereas they expect women to do the female-assigned tasks (i.e., laundry or cleaning), which could lead them to evaluate a given division of housework as appropriate even though they expect gendered housework tasks (e.g., Schneider, 2012). Therefore, including information on the specific housework tasks might be better suited to measuring the gender deviance neutralization norm.

Second, recent research has introduced a new dimension of unpaid work, the so-called "mental load", for which women are predominately responsible (e.g, Daminger, 2019; Dean et al., 2022; Ruppanner et al., 2021). Mental load is a combination of cognitive and emotional work that consists of being responsible for everyday tasks that are often invisible, such as organizing appointments and/or dividing tasks within the family (see e.g, Daminger, 2019). If respondents assumed women's responsibility for the mental load within a couple while answering the vignettes, this would bias the results. One could argue that as information was provided on the partners' shares of housework and child care, it is likely that respondents simply assumed that the person responsible for these tasks also assume the associated mental load. Still, considering mental load as another factor of couples' division of paid and unpaid work is an interesting starting

[27] Depending on whether or not information on labor market characteristics is provided, the factorial survey experiment on couples' division of work contains only four, six, or seven dimensions. Arguably, it might be challenging for respondents who apply equity norms to count the different labor market characteristics against each other. Still, even if some respondents had difficulties with this, it should not challenge the conclusion that the gender of the vignette person does not play a role in the evaluations.

point for new research on gender norms.

Third, newer research by Grunow et al. (2018) and Knight and Brinton (2017) has shown that gender norms are not to be interpreted as ranging unidimensionally from "traditional" to "egalitarian", but rather as a multidimensional construct. They argue that individuals might support an equal labor market participation of men and women and simultaneously expect women to be mainly responsible for child care. A factorial survey experiment is well-suited to measuring multidimensional constructs, as it allows relevant dimensions to be varied independently from one another (Auspurg and Hinz, 2015). In chapter 3, the findings hint at a normative expectation that women should take on slightly more child care than their male partners, while at the same time doing an equal amount of paid work. However, these effects are very small and not always visible and should therefore not be over-interpreted. This dissertation mainly focuses on couples' division of housework and paid work, so the focus of the factorial survey experiment used is not predominantly on gender norms in regard to responsibility for child care. For example, there is no information included as to how the child feels or rather if it suffers due to the parents' division of child care. Therefore, newer research should focus more on normative judgments of parents' division of paid work and child care.

Finally, this dissertation focuses only on gender norms regarding the division of paid and unpaid work of heterosexually partnered men and women. Gender norms of non-heterosexual couples or individuals that identify as trans or nonbinary were not examined. The intersection of racial and gender norms was also not considered here, although this is often cited as the golden standard in research on attitudes, norms, and discrimination (e.g., Scarborough et al., 2021). Research has been published on the division of paid and unpaid work among homosexual couples (e.g., Kelly and Hauck, 2015; Bauer, 2016), on interaction effects of individuals' gender and race on internalized gender norms (e.g., Scott Carter et al., 2009), and on social norms regarding gender and race (e.g., Scarborough et al., 2021). A factorial survey experiment might best combine all directions of research to provide broader insights on gender-specific social norms.

1.7.4 Outlook

This dissertation found that it is widely believed that men's and women's (relative) participation in the labor market should influence their time in housework, not their gender. This finding not only questions the existence of an unconditional gender norm, but questions the overall existence of a gender-specific norm in context of division of labor. What does this imply for future research? As of now, theories on gender norms have been heavily used to explain the gendered division of work. Typically, it has been argued that egalitarian individuals divide their paid and unpaid work more equally than

traditional individuals. This hypothesis has often been confirmed using the standard item measurement of gender norms. However, if it is assumed that individuals do not actually differ in their normative belief on who should be responsible for housework (i.e., a prescriptive gender norm), but rather in their belief on who is actually responsible for housework (i.e., a descriptive gender norm), the impact of gender norms on individual behavior has likely been overestimated.[28]

If it is not (prescriptive) gender norms, what does influence couples' division of paid and unpaid work? This question cannot be answered based on the results presented here. Previous research has argued that individuals' gender identities or rather, self-classification as men or women (measured either through sex-differentiated interests or self-categorization), leads them to assume different roles their genders are typically responsible for (e.g., Wood and Eagly, 2015). However, such an identification as men or women would lead to a gendered division of work only in combination with societal expectations (Wood and Eagly, 2015, p. 466). Moreover, it is questionable whether housework can (still) be used as a gender identity measure, as there is evidence that men and women do not (anymore) differ in their preference for and consequently interest in doing housework (Auspurg et al., 2017b). It is conceivable, however, that men and women still do or not do child care as a way of expressing their gender identity.

Others argue that women's skills for unpaid work are greater than men's; for example, that women have a biological advantage in child care (Becker, 1993). This view is challenged by the argument that women's advantage in child care is only socially constructed and that men are equally capable of taking care of a child (e.g., Gregory and Milner, 2011; Johnston and Swanson, 2006; Murray, 1996). Regardless if this advantage has been socialized or biologically predefined, both could be cause for a gender gap in child care and also in housework, especially if women anticipate their responsibilities for unpaid work and therefore accumulate less human capital than men (Becker, 1993). In addition, employers' expectation that women are mainly responsible for housework and child care might lead to statistical discrimination of women, and hence, to lower earnings (Arrow, 1973; Phelps, 1972). As a result, women have less human capital, earn relatively less, and spend less time in the labor market than their male partners. Based on the economic theories, this would lead to women being mainly responsible for unpaid work (see, e.g., Becker, 1993; Blood and Wolfe, 1960; Ott, 1992; Presser, 1994), even though there is no gender-specific social norm.

The big question as to why women do more of the unpaid work than men cannot be answered with this dissertation. At the very least, I find experimental evidence for the existence of a norm that couples should not base their division of paid and unpaid

[28] This naturally leads to the question as to why individuals differ in their descriptive gender norms, which is left to future research. However, one possible mechanism could be that individuals base their descriptive gender norms ("what is most likely the case") on their personal experiences of the world. If these experiences differ, descriptive gender norms would also differ.

work on gender, but on relative income and time availability.

Appendix

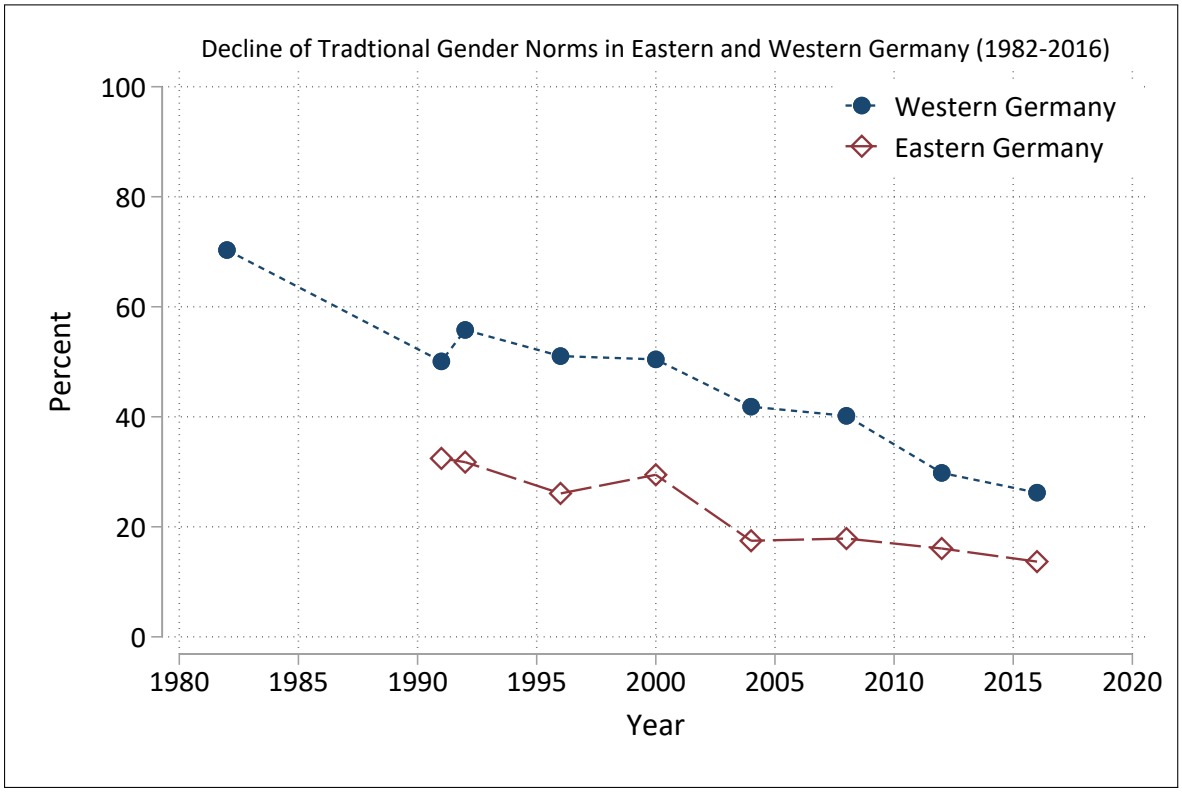


Figure 1.A1. Decline of Traditional Gender Norms in Eastern and Western Germany (1982-2016)

Source: ALLBUS 1982-2018 (GESIS - Leibniz-Institut für Sozialwissenschaften, 2021). For 2012 and 2016 only split A was used. Own calculations.

2

Probing on the Conditionality of Gender Norms: Is the Significance of Gender Overrated?

(Katrin Auspurg and Sabine Düval)

Unpublished Article.

2 Probing on the Conditionality of Gender Norms: Is the Significance of Gender Overrated?

Abstract Gender inequalities in housework persist in all Western societies, and gender norms are considered an important predictor of the division of paid and unpaid work among couples. In this article, we examine the validity of widely used item questions on gender norms and argue that existing research does not consider the conditionality of (gender) norms. However, if gender norms are indeed conditionally applied, past research using standard measures has likely overestimated the genuine role of gender as an evaluation criterion. A total of 1,232 German residents participated in a multifactorial survey experiment on the appropriate division of housework of hypothetical heterosexual couples in 2017/2018. In these vignettes, men's and women's labor market characteristics (i.e., income and labor market hours), division of housework, and family status were varied experimentally across descriptions. In addition, the amount of information presented to the respondents was randomly varied. This experimental design allowed for the distinction of the genuine effect of gender from other competing explanations, such as equity norms, by conducting a causal mediation analysis. We find that for "traditional" respondents (defined via item question), women are expected to do more housework than men only in the absence of information on labor market characteristics. Once this information is available, even these respondents report that the appropriate share of housework depends on gender-neutral factors such as monetary contributions and time availability. Overall, equity norms appear to trump gender norms.

2.1 Introduction

In this article, we demonstrate that standard item measurements of gender norms or gender role attitudes likely overestimate the genuine role of gender as an evaluation criterion. Using a more nuanced, novel experimental design, we find broad evidence for equity norms as opposed to gender norms: The appropriate amount of housework is seen to strongly depend on individuals' contributions to the partnership in terms of relative earnings and labor market hours, with hardly any impact of gender.

Significant gender inequalities in paid and unpaid work persist in all Western countries. Studies on couples in the U.S. and Germany indicate that women (still) have lower employment rates and relative earnings than men (e.g., Killewald and Gough, 2010; Nitsche and Grunow, 2016). A partial explanation is given by women's continuing responsibility for the majority of child care and housework. Official statistics indicate that on average, women are responsible for roughly two-thirds of the housework in heterosexual couples (e.g., German Federal Statistical Office, 2015). One often discussed explanation for this gender differences in paid and unpaid work are gender norms, often also called gender ideologies or gender role attitudes. Gender norms are social expectations of appropriate behavior for men and women that are attributed to them based solely on their gender (e.g., Bartley et al., 2005; Davis and Greenstein, 2009). Individuals may internalize these prescriptive norms during the socialization process and/or they may act in accordance to these norms to avoid social sanctions or to clearly display their gender identities (Akerlof and Kranton, 2000; West and Zimmerman, 1987).

The standard approach to measuring gender norms is to ask respondents to evaluate short item statements that represent work divisions within couples, such as "A man's job is to earn money; a woman's job is to look after the home and family." Other statements ask about gendered career preferences or competencies of mothers versus fathers. According to response patterns, respondents are then classified on a summated scale as being more or less supportive of traditional gender norms (for an overview on frequently used items, see, e.g., Walter, 2018a). Dozens of studies suggest that the apparent gender norms measured in this way can predict the share of (paid and unpaid) work within couples, as well as variations in work arrangements that exist over the life course or across different household types (e.g., married vs. unmarried couples; parents vs. non-parents; for overviews, see, e.g., Coltrane, 2000; Davis and Greenstein, 2009). Gender norms have also been found to explain significant parts of cross-country differences in gender inequalities (Fortin, 2005; Nordenmark, 2004).

However, this paper argues that the standard item measurements have strong limitations that have thus far been overseen in previous literature. First, unidimensional item questions do not allow for the study of the conditionality of gender norms: Are

female partners strictly expected to do a larger share of housework than their male partners? Or does this only hold under specific conditions, such as women having lower earnings power or more time available than men? Second, when gender norms are conditional, respondents must impute missing information about the proposed conditions answering abstract questions. The most plausible assumption is a reliance on stereotypical beliefs about men and women in general. In this case, they may agree with traditional gender roles simply because they assume that women have lower labor force participation and/or earning power than men. If so, equity would then be the reasoning behind women doing more housework, not gender norms.

This paper probes the conditionality of gender norms and examines the mechanisms responsible for agreement with traditional item questions. The methodological approach is based on Acharya et al. (2018), who propose the use of multifactorial survey designs to disentangle different mechanisms by way of causal mediation analysis. More precisely, they suggest to not only randomly vary the treatment variable across respondents, but to also randomly provide or withhold information on relevant mediating variables (Acharya et al., 2018). This paper therefore uses a multifactorial survey experiment in which respondents were asked to evaluate a set of hypothetical work-family arrangements. Across the descriptions, the vignette person's gender (i.e., the treatment variable) is independently varied. In addition, both partners' labor market hours, relative labor market income, and further characteristics were randomly varied (i.e., the mediating variables). In order to apply mediation analysis, three split samples were included in which more or less information on the partners' relative earnings and time availability (i.e., assumed mediating variables) were presented. This design allows for an experimental probe into whether gender norms are the true reason respondents agree to traditional housework arrangements (i.e., women are responsible for a larger share of household tasks). The alternative explanation is that respondents impute missing information in a gender-specific way (i.e., assuming gender inequalities in paid work) and therefore agree to traditional housework arrangements when, in fact, they follow nontraditional equity or equality norms.

This experiment was embedded in the 10th wave of the German family panel (Brüderl et al., 2019), a large scale population survey in Germany collected in 2017/18. A total of 3,693 evaluations of work-family scenarios were gathered from 1,232 respondents. With this rich experimental evidence, support for the conditionality of gender norms is clear: Respondents support arrangements in which the female partner shoulders a majority of the housework only when there are gender inequalities in other realms, such as labor force participation and earning power. These results also clearly indicate that the standard item measurements—that overlook this conditionality—overestimate the prevalence of gender norms. In sum, these results suggest that existing literature might have overestimated the impact of gender norms on actual work arrangements within

couples.

2.2 The State of the Art

2.2.1 Gender Inequalities Explained by Gender Norms

One relevant explanation of gender inequalities found by previous research lies in the societal expectations of appropriate behavior attributed to men and women based solely on their gender (see, e.g., Bartley et al., 2005; Davis and Greenstein, 2009).[29] More specifically, it is assumed that men and women base their shares of paid and unpaid work on so-called gender norms. Terms such as "gender ideologies" or "gender role attitudes" have also been used (mostly interchangeably) to refer to the same underlying idea.[30]

The existence of prescriptive gender norms has to a large extent been rationalized by the fact that they allow for a more efficient specialization and aid in the avoidance of conflicts over the division of paid and unpaid work by pre-specifying separate spheres for men and women (Thibaut and Kelley, 1959). Individuals may internalize these norms during the socialization process and other priming experiences. In this case, expectations concerning individual behavior become the ends that individual's desire (Elster, 1989; Horne, 2003).[31] Gender norms may also function as a lens or frame to evaluate one's own and others' behavior, as they may define reference points for expectations of typical or fair contributions to (house)work within couples (Greenstein, 1996; Ridgeway, 2011).

Numerous studies have consistently found gender norms to be an important predictor of couples' division of paid and unpaid work (for overviews, see, e.g., Coltrane, 2000; Davis and Greenstein, 2009). Gender norms were also found to explain a substantial part of cross-country differences in housework patterns and in the level of female involvement in paid work (Nordenmark, 2004). Similarly, time trends in gender inequalities over the course of relationships have been attributed to changes in gender

[29] Other well-known explanations based on economic exchange and bargaining theories assume gender inequalities to be the result of differences in men's and women's labor market resources, earning power, and available time (see, e.g., Becker, 1985, 1993; Blood and Wolfe, 1960; Presser, 1994).

[30] Other terms include "gender roles," "beliefs about gender," "attitudes about gender," "gender-related attitudes," "gender egalitarianism," and many more (see Davis and Greenstein, 2009). We suggest the general use of the concept of norms, which allows us to draw on the rich theoretical and empirical literature to distinguish different norms and to explain when and why norms emerge and erode (see, e.g., Opp, 2001). Note that the following methodological arguments also apply to the more vague concepts of gender ideologies and gender role attitudes.

[31] There is no external force needed to push individuals to conform to these expectations (Bittman et al., 2003). It could be, however, that individuals have only partly internalized these norms and follow them mostly to avoid social sanctions or to receive social approval by demonstrating competence in "doing gender" (West and Zimmerman, 1987). In doing so, individuals can create and signal their identity as men and women (West and Zimmerman, 1987; Akerlof and Kranton, 2000).

norms (Nitsche and Grunow, 2016). Moreover, gender norms were shown to be an important source of differences in work arrangements between married and cohabiting couples, as well as parents and non-parents (see, e.g., Baxter et al., 2008; Gupta, 1999). It is likely that at least some of these differences could have stemmed from a selection bias of individuals with more traditional gender norms into marriage and parenthood. Still, a large part of the existing gender differences in paid and unpaid work is attributed to gender norms as one of the main predictors of gender inequalities.

2.2.2 On the Measurement of Gender Norms

Probing on the Conditionality of Gender Norms

In survey research, gender norms are most commonly measured with a summated rating scale. Respondents are asked to evaluate different statements regarding the appropriate behavior of men and women in (heterosexual) intimate relationships (sometimes also only one item question is used; for overviews: Davis and Greenstein, 2009; Walter, 2018a). Examples of standard items include "A man's job is to earn money; a women's job is to look after the home and family," (ISSP, 2021) and "All in all, family life suffers when the woman has a full-time job" (ISSP, 2021).

This approach to measuring gender norms has often been criticized (for a recent overview: Walter, 2018a); the key issues are mentioned here. First, most items only portray traditional work arrangements, and therefore respondents do not have sufficient opportunities to express egalitarian views (Braun, 2008). Second, items frequently focus only on the role of women, while neglecting the roles of men and particularly fathers. Third, due to attitudinal change the item scales designed in the 1960s and 1970s no longer provide enough variance to discriminate between respondents' attitudes: When using the classic items today, nearly all respondents are classified as egalitarian, i.e., not supporting any gender norms. Fourth, recent research has found that gender norms no longer appear on a unidimensional scale ranging from "traditional" (i.e., preference for separate spheres) to "egalitarian" (i.e., preference for joint spheres) (see, e.g., Cotter et al., 2011; Grunow et al., 2018). Instead, multidimensional concepts (i.e., support for either joint or separate earning and caring, and at the same time either favoritism of individual choice or predefined gendered traits) are widespread in eight European countries (Grunow et al., 2018).[32] Based on this framework, Grunow et al. (2018) identified five different profiles: egalitarian, egalitarian essentialism, intensive parenting, moderate traditional, and traditional.[33] The most prevalent are egalitarian gender norms, whereas traditional gender norms were found to be rare. A multifactorial survey

[32]The countries are the Czech Republic, Poland, Western Germany, Switzerland, Italy, Spain, the Netherlands, and Sweden (Grunow et al., 2018).

[33]For a comprehensive comparison of these different manifestations of gender norms, see Grunow et al. (2018).

experiment is well-suited to account for the multi-dimensionality of gender norms, as it allows for the independent variation of the number of dimensions in order to measure their impact on respondents' evaluations.

An important feature of social norms is their conditionality (e.g., Bicchieri, 2010; Bicchieri and Chavez, 2010, 2013; Elster, 2011; Horne, 2003; Jasso and Opp, 1997; Kahnemann et al., 1986; Opp, 2020, and many others). This article discusses the major shortcoming of previous research ignoring the conditionality of gender norms. We argue that a measurement instrument that does not capture the conditionality of gender norms likely overestimates the importance of gender norms for the division of paid and unpaid work among couples.

Most social norms are not generic imperatives, but only hold under specific conditions.[34] This conditionality can refer to both characteristics of the social situation and/or characteristics of the protagonists (e.g., Diefenbach and Opp, 2007; Horne, 2003; Jasso and Opp, 1997, p. 948; and many others). One reason for the conditionality of norms are competing norms: In many social situations, several, often conflicting norms apply in parallel, requiring individuals to trade-off different normative expectations. For instance, there may be competing norms on the adequate financial provision of families (such as families being able to live without social benefits) or norms on appropriate child care (such as both parents being expected to spend time with their children). Conflicting normative expectations might dampen or also intensify (gender) norms. The latter presumption is implied in "gender display" or "gender deviance neutralization" theories that assume that women will hold true to traditional household responsibilities if they have a higher labor market income than their male partners (Schneider, 2011).[35] Fulfilling traditional gender housework norms then allows for the female partner to compensate for her violation of the gender norm for paid labor (Akerlof and Kranton, 2000; West and Zimmerman, 1987).

Furthermore, not only expectations based on gender norms might be in conflict in some situations. Alternative social norms on the appropriate distribution of work can be found in the literature on distributive justice (Deutsch, 1975). A first principle is the equity norm, fulfilled when both partners' rewards (e.g., leisure time) are distributed proportionally to their input to the relationship (e.g., labor market and housework hours) (Gager, 1998; Thompson, 1991). Following this equity norm, the higher an individual's labor force participation relative to his/her partner's, the more he/she should be entitled to reduce housework duties. Monetary inputs may also be of importance, meaning that a higher relative contribution to the household income, all else equal, may entitle individuals to reduce their housework chores. A second norm assumed to

[34] Even for the norm "you must not kill" there are conditions for which the norm does not apply, e.g., in wartime or in self-defense (Jasso and Opp, 1997).

[35] Women's employment might be needed to financially support the family, for example, because the men's earnings are be too low.

apply to the division of housework is the norm of equality (see Jasso, 1983). It defines a situation in which all rewards and duties are split equally (50:50) between the partners as appropriate. Both men and women are expected to do half of the housework chores and also expected to contribute to the same extent to the household income.

These alternative norms prescribe that housework is allocated gender-neutrally and solely based on individuals' input to the relationship. Therefore, the separation of gender norms from competing norms boils down to the question of whether there is an effect of gender in itself—only when a genuine gender effect is apparent can we speak of gender norms. Regarding housework, the gender effect should exist with women (as opposed to their male partners) being held responsible for housework and child care, even with similar or even greater labor market participation than her male partner ("gender trumps money;" Bittman et al., 2003; Brines, 1993, 1994).

Have Gender Ideologies Been Overrated Due to Measurement Limitations?
Typical questions used to measure prescriptive gender norms include very broad statements such as "Housework should be equally split between men and women." Respondents generally evaluate these and further statements on a five-point scale from "do not agree at all" to "fully agree." For such global questions, respondents can only make unconditional statements, as they are not able to differentiate their judgments depending on the respective social situation (e.g., male vs. female partner's earning power). It remains unclear whether and how the support of this norm depends on any social circumstances the couple is facing, and hence which impact gender has in relation to any other characteristics of the two partners. At first glance, this may not seem problematic for many applications. Many studies are not interested in measuring effect heterogeneity, but only in identifying respondents who support—on average—gender norms to a different degree. However, when questions are posed without social context, the items do not allow for the differentiation between descriptive and prescriptive norms. Even worse, they likely confuse gender and equity norms. In the case that gender norms are applied conditionally, respondents must impute the missing information about social conditions themselves. Respondents are very likely to fall back on stereotypical beliefs about the "typical" division of labor, or use their own personal experience as a reference.

In all Western countries, the "typical" scenario involves the male partner working more hours in a paid labor, therefore contributing more to the household income than his female partner (OECD, 2017). In Germany, the most typical constellation involves the male partner employed full-time while the female partner works part-time, with women's hourly pay on average 18% percent lower than that of men (Gartner and Hinz, 2009; German Federal Statistical Office, 2022a; Nitsche and Grunow, 2016). If respondents imagine this type of couple when responding to item questions, they might

agree with traditional housework arrangements even if they themselves do not adhere to gender norms at all, simply because they follow norms of equity. In this case, the rationale is that the housework should not be equally split but mainly performed by the female partner, as she is assumed to have less labor market hours, therefore contributing less to the household income, and that she therefore has more time for housework.[36]

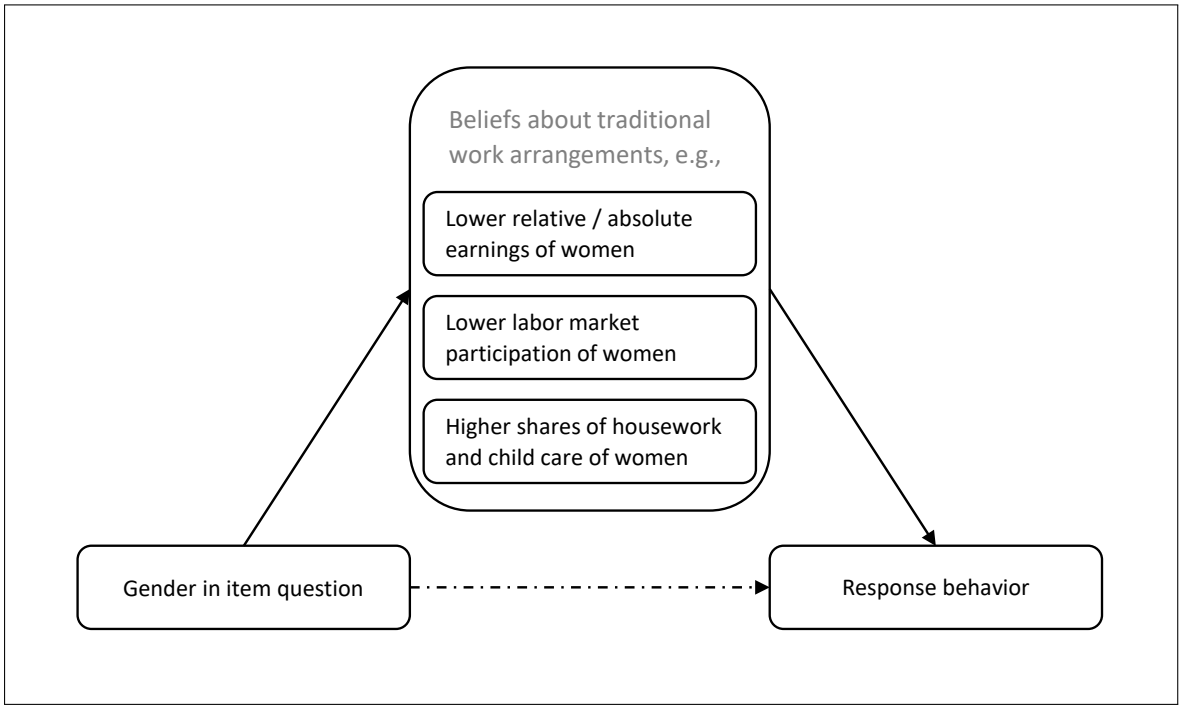


Figure 2.1. Effect of Gender Presented in the Item Question on Response Behavior

Note: This figure shows the assumed causality underlying most literature using standard items to measure gender norms. The response behavior is attributed to the gender presented in the item question (dotted arrow). We argue, however, that this correlation could be mediated by respondents' stereotypical beliefs about traditional work arrangements. These beliefs could—together with equity norms—give a strong impetus to agree with traditionally worded item questions, even when respondents do not adhere to gender norms (solid line). This makes it impossible to distinguish whether respondents agree to traditional item questions based on gender norms or equity norms.
Source: Own illustration.

This argument is illustrated in Figure 2.1. The dotted arrow shows the standard assumption. Respondents' answering behavior is attributed to the gender represented in the item question. Ultimately, the item questions are interpreted to mean that changing the gender of the item question from female to male changes the response behavior. For example, traditional respondents are expected to agree (disagree) with the statement that women (men) should do the housework. The solid arrows show

[36] Meanwhile, items that include some information on the labor market status (e.g., one item in the NSFH (1988) was: "If a husband and wife both work full time, they should share household tasks equally") are sometimes implemented. Even so, given the actual gap in hourly pay, respondents may assume that the man out-earns his wife, which, given equity norms, may still entitle him to do less housework without gender norms being applied.

our alternative assumption: The gender presented in the item question could prime individuals' beliefs about existing work arrangements by gender. These (descriptive) beliefs could—with the support of equity norms—give a strong impetus to agree with statements on traditional divisions of housework, even when individuals do not adhere to gender norms themselves. If this alternative pathway holds, most standard measures have likely overestimated the prevalence of gender norms. To distinguish whether respondents actually support gender norms or if their (traditional) response behavior is caused by (descriptive) beliefs about traditional work arrangements, the two pathways must be disentangled.

Often, the observation is made that respondents share in particular traditional gender norms after the transition to parenthood (e.g., Katz-Wise et al., 2010; Perales et al., 2015). This is typically rationalized by the fact that parenthood makes norms on the adequate behavior of parents more salient. If the assumptions presented here are correct, it could, however, also be true that parents consider gendered labor market inequalities that are especially high for parents (due to a "fatherhood wage premium" vs. "motherhood wage penalty"; or many mothers no longer participating in the labor market or reducing their labor market hours). In addition, time trends might be seen in a different light: perhaps it is not declining gender norms that lead to a decline in traditional work patterns, but rather the other way round. Evidence from trend studies shows that declines in men's status as a breadwinner or changes in statutory parental leave arrangements can have norm-setting effects by subsequently altering individuals' agreement with gender norms (see, e.g., Gangl and Ziefle, 2015; Zuo, 2004). However, to the best of our knowledge, the idea that existing gender labor market inequalities may also form a lens through which the item questions on gender norms are interpreted has thus far not been well reflected in previous literature.

2.2.3 Insights by Experimental Designs

To understand the actual causes of gender inequalities, the various explanatory mechanisms must be taken apart. To achieve this, a measure that can isolate the direct effect of gender is necessary: Does gender on its own have an effect? In other words, is agreement with traditional statements about couples' division of housework based on gender norms? Or is the effect of gender eliminated when relevant mediators, e.g., men's and women's labor market resources (e.g., relative earnings and employment hours) are controlled for, as supposed by norms of equity? To see to what extent individuals follow gender versus equity norms, the presumed conditionality of norms would ideally be investigated. In methodological terms, the total effect of gender must be decomposed into a direct effect (the effect of gender in itself) and an indirect effect (the part of the effect mediated through beliefs about typical work-family arrangements) to disentangle the various explanatory approaches. This approach calls for causal mediation analysis,

best executed with experimental designs that allow for the manipulation of relevant mediator variables (Acharya et al., 2018).

Multifactorial survey experiments have long been used as an ideal tool to measure attitudes and social norms (see Jasso and Opp, 1997; Opp, 2002, 2020). In recent research, factorial survey designs have also been used to show that preferences about ideal work arrangements strongly depend on situational circumstances, such as individuals' access to resources and institutional constraints (see Auspurg et al., 2017b; Jacobs and Gerson, 2016; Pedulla and Thébaud, 2015). Other experimental studies have demonstrated that individuals hold gendered descriptive norms and status beliefs (Auspurg et al., 2017a; Correll et al., 2007). In addition, some experimental studies have focused on different housework standards men and women are held responsible for, or on fairness norms regarding the division of work (Carriero and Todesco, 2017; Schulz, 2021; Thébaud et al., 2021).[37] To the best of our knowledge, no study has explicitly focused on prescriptive gender norms (i.e., "what should men and women do?") in housework division.

This is the first study to use a multifactorial survey experiment to probe the conditionality of gender norms. The experimental approach is based on a design proposed by Acharya et al. (2018), which allows the direct effect of gender (i.e., the effect of gender in itself) to be distinguished from indirect effects of gender, which are mediated through, for example, beliefs about typical work-family arrangements. This approach enables the explicit examination of the conditionality of gender norms, as well as a direct test of the assumed limitations of the standard item measurements: Do (traditional) respondents simply agree with traditional work arrangements because they are lacking information on couples' labor market resources and, hence, assume typical gendered constellations? Several authors of recent review articles have suggested the use of multifactorial experimental methods to measure (gender) norms in the context of the division of housework (e.g., Bicchieri et al., 2014; Davis and Greenstein, 2009; Mutz, 2011; Thébaud et al., 2021), a line of research this study falls in line with.

2.3 Research Design

2.3.1 Our Experimental Setup

In this multifactorial survey experiment, respondents were presented with short texts containing information about work-family scenarios of hypothetical heterosexual couples. Each respondent was asked to rate the appropriateness of a vignette person's individual share of the couple's joint housework in three different vignettes. The vignettes consisted of several attributes (so-called dimensions) that were randomly varied in their levels. The possible ratings range on an eleven-point scale from -5 "Her

[37] More information on existing experimental studies is provided in Appendix B.

/ His housework share should be much smaller" over 0 "Her / His housework share is appropriate" to +5 "Her / His housework share should be much larger". Note that randomization was used between respondents, meaning that either the proportion of housework done by a female or a male vignette person was to be rated.[38] Respondents were also randomly assigned to one of three splits, between which the amount of information on labor market characteristics was also varied. This resulted in conditions with low information (i.e., no information on any labor market characteristics), medium information (only information on labor market hours), and high information (information on labor market hours and relative income). For more information on the design, see Appendix B. For an example vignette, see Figure 2.B1 in Appendix B.

This experimental design enables the application of causal mediation analysis in a survey experiment, as proposed by Acharya et al. (2018). In a first step, vignettes of the low information condition were used, with no information on the labor market characteristics of the partners, which are similar to standard item questions. As the gender of the vignette actor was randomly varied between respondents, we suspect an effect of the vignette person's gender on the appropriateness of his or her share of the couple's housework, especially so for "traditional"[39] respondents. This effect has been historically seen as proof for traditional gender norms; in other words, when respondents agree with a traditional item, they are ascribed to a traditional gender norm. Our argument, however, is that even if no information is provided on labor market characteristics (i.e., the mediator variable), it will still have an effect on the outcome: The mediator takes its natural value, meaning the value it would take under the assigned treatment condition (i.e., the assigned gender of the vignette person). If respondents are to evaluate the housework share of a woman, they will then impute stereotypical descriptive beliefs about lower labor market hours and income. Likewise, a man would be expected to earn and work more. The low information experimental split (and standard item questions) make it impossible to differentiate whether a gender effect is due to the (random) variation of gender, or due to different labor market characteristics attributed by gender.

In a second step, the gender of the vignette person was again randomized between respondents. Further, randomized information was provided on 1) the vignette person's labor market hours or 2) the vignette person's labor market hours and relative income by way of vignettes from the medium and high information splits, respectively (see Table 2.B1 in Appendix B). Hence, the mediator no longer depends on the assigned (gender) treatment. This two-step approach allows the research question to be illuminated: Does an effect of gender remain for traditional respondents even after extensive

[38] A between-respondent split was used deliberately to avoid social desirability bias, as well as confusion when respondents switch between the evaluation of men's and women's household shares.

[39] "Traditional" as measured by the standard item questions on gender norms.

information on labor market characteristics is provided? If so, this points toward the relevance of gender norms: Women are—despite the same labor market power as men (in the vignette universe)—expected to do a larger share of the housework. However, if the effect of gender disappears when information on labor market characteristics is provided, this indicates that given the same resources, men and women are expected to do the same amounts of housework, supporting equity norms.

2.3.2 Respondent Sample and Questionnaire

This multifactorial survey experiment was implemented in wave 10 of the large-scale German Family Panel (pairfam) in 2017/18. A detailed description of the study can be found in Huinink et al. (2011). Note that the pairfam survey is based on a random sample of three birth cohorts of the German population. External validity of this experiment is much higher than that of the commonly used samples of college students or Amazon MTurk. Germany is also particularly suited for this experiment as cross-country research has revealed that a relatively high proportion of Germans share traditional gender norms (OECD, 2017), at least when measured with standard item questions. In addition, Germany is a conservative welfare state that shows relatively pronounced gender inequalities, for example in terms of large gender wage gaps and relatively low female labor market participation (e.g., Gartner and Hinz, 2009; German Federal Statistical Office, 2022a). Both aspects made Germany an ideal context for testing the assumption that part of the gender norms measured by standard items simply reflect respondents' stereotypical assumptions about gender inequalities as they actually exist.

Of the 4,750 respondents that participated in wave 10 of the pairfam panel, a randomly selected subsample of 1,283 respondents was allocated to this experiment.[40] All observations with missing values on relevant items concerning individuals' gender norms measured by standard item questions, respondent age, and respondent gender were excluded, as well as all evaluations of respondents that were in a homosexual relationship at the time of the interview.[41] The final sample size consisted of 3,693 vignette evaluations from 1,232 respondents.

[40] The other respondents participated in a different experiment, in which they were asked not to evaluate the division of housework sketched in the vignettes, but instead the work division in general. These experiments are referenced in section 2.4.2

[41] This was done as these respondents probably adhere specific gender norms, but not enough cases were available to run separate analyses on this subsample. Respondents not living in a relationship were kept in the analyses, but the main results presented here do also hold when we exclude this sample.

2.3.3 Estimation Strategy

The following analyses follow the standard approach in factorial survey research (Auspurg and Hinz, 2015; Hox et al., 1991) and use linear regressions to predict vignette evaluations. The first step is to estimate regression models with random intercepts, followed by robustness checks with alternative model specifications (see section 2.4.2). Dummy-variables are included for the different levels of each vignette dimension. As stated, first only vignettes with the low information condition are used. In a second step, vignettes of the medium and high information conditions are used.

The dependent variable is the evaluation of the share of housework done by the female or male vignette person. To allow for pooled analyses on both genders, one unified response scale was generated to indicate the evaluation of the female vignette person's housework share. The resulting evaluation allows for simplified interpretations: positive (negative) values indicate that the female vignette person should do more (less) housework than indicated in the vignette, and a zero effect indicates that the housework share is considered appropriate. In addition to the factorial survey experiment, all respondents answered a standard item battery on gender norms. These analyses focus on the one item closest to the vignette scenarios with low information: "Men should participate in housework to the same extent as women." Respondents could answer on a 5-point scale ranging from 1 "Disagree completely" to 5 "Agree completely". Respondents that answered 1 or 2 are classified as traditional, those that answered 4 or 5 as egalitarian, while the others (answer 3) are neither nor. Whether the two groups not classified as egalitarian by this measure (roughly 20% of the sample) also shift to an egalitarian view once provided with more information is investigated in the alternative multifactorial vignette measurement.

To ease interpretation, most results will be presented visually in the following. These graphs plot the point estimates of regression coefficients together with their confidence intervals. Regression tables will full information on the underlying model estimates can be found in Appendix C. Descriptive statistics and additional analyses can be found in Appendix A.

2.4 Results

2.4.1 Main Results

Figure 2.2 shows the impact of the vignette person being displayed as a woman (vs. a man) on the appropriateness of her (his) share of housework. Results are shown separately for three groups of respondents based on their answers to the standard item questions: Respondents' gender norms are either classified as "traditional," "egalitarian," or "neither nor." In each graph, results are depicted separately by the amount

of information provided on the couple's labor market characteristics. "No info" here indicates that neither information on labor market hours nor relative income was presented. In the other two information conditions, information on labor market hours (medium information) or labor market hours and relative income (high information) were included in the vignettes.

In the "no info" condition, as expected, only respondents categorized as "traditional" by the standard item questions rate that women should do more housework compared to their male partners, all else held equal. Here, the item and vignette measurements show a high level of reliability. However, the responses to the vignettes provide valuable additional information: The amount of hours women should invest more can be calculated using cross-elasticities between the gender of the vignette person and indicated housework hours.[42] On average, "traditional" respondents' ratings indicate that women are expected to engage in 10 hours more housework than men. Given the vignette specifications, this means that women should do 20 hours (66%) of housework per week while men should do 10 hours (33%), which is close to the actual mean share of housework done by women in Germany (about 60%, see German Federal Statistical Office, 2015). Adding the information that the female partner has, on average, the same labor market hours (and income) as the male partner removes the significance of the gender effect, even for "traditional" respondents. In other words, this group only supports a gendered division of housework when they lack information on the labor market status of the couple.

For "egalitarian" respondents, the effect turns slightly negative, meaning that women who are described as having the same labor market characteristics as their male partners are seen as being entitled to do slightly less housework. Following our estimates, "egalitarian" respondents find it appropriate for the female partner to engage in 1.7 hours less than their male partners, or roughly 47% of the overall housework.

In sum, adding information shifts the appropriate share of housework for women to lesser amounts. The most plausible explanation for this is that higher labor market hours and financial contributions entitle individuals to do less housework (as the equity norm suggests). As expected, without explicit information, "traditional" respondents assume that women have less of these labor market resources than their male partners, as would also be the standard case in Germany.

As with the design itself, the levels of these labor market resources were experimentally varied to gain deeper insights on whether the interpretations are correct. Figure 2.3 is therefore quite similar to Figure 2.2; however, the focus shifts to the effect of labor market hours in more detail. The first line of each of the three graphs indicated

[42] These predictions were estimated by means of cross-elasticities, making use of the Stata ado "wtp" (Auspurg and Hinz, 2015; Hole, 2007). The logic is to calculate the amount by which the share of housework as indicated in a vignette would need to be increased (or decreased) to bring the evaluation down to zero. For details and formulas, see Auspurg and Hinz (2015).

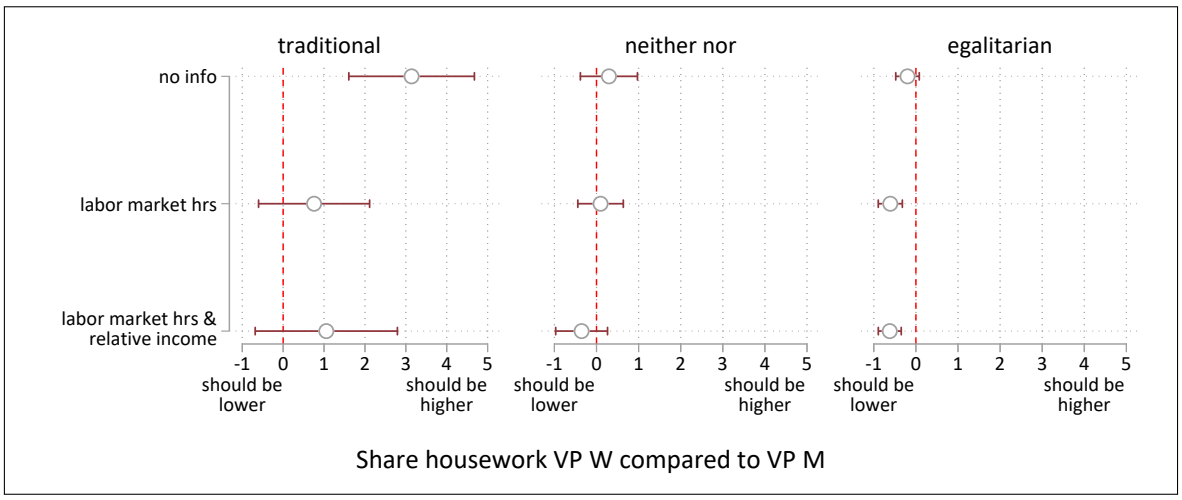


Figure 2.2. Effect of Adding Information on Labor Market Characteristics by Respondent Group

Note: This figure shows the marginal effect of adding information on labor market characteristics ("no info," "labor market hours," and "labor market hours and relative income") on the appropriateness of the female vignette person's share of housework compared to the male vignette person's. Effects are shown separately for "traditional" (first column), "neither nor" (second column), and "egalitarian" (last column) respondents. Results were estimated separately by the respondents' gender norms measured by the item question "Men should participate in housework to the same extent as women." The corresponding regression tables can be found in Table 2.C1 in Appendix C.
Number of vignette evaluations of "traditional" respondents: 117; "neither nor" respondents: 647; "egalitarian" respondents: 2,929.

the effect of the "no info" condition. The further lines show the effect for a woman conditional on her different relative labor market hours (ranging from 20 hours more to 20 hours less than her partner).

First, it is clear that labor market hours have a strong impact in the direction predicted by equity theories: The higher the relative share of labor market hours, the lower the appropriate share of housework. Second, these equity norms are shared by all three respondent groups. Third, and even more interesting: For which level of labor market hours is the effect of being a woman on par with the effect of the no information condition? The former is evidence that respondents supposed a similar labor market level when explicit information on this dimension was excluded from the vignette. In this regard, the most substantial difference across the three groups of respondents can be observed: For both "egalitarian" and "neither nor" respondents, the effect of the "both same hours" level matches the effect of the "no info" condition. This suggests that these respondents, on average, expected the female vignette person to have a similar labor market participation as her partner. Furthermore, these two groups rated a division of work appropriate for which both partners do the same amount of housework, given their labor market characteristics are the same.

The pattern found in the "traditional" respondents is different: For this group, the plotted effects do not increase as clearly with the relative labor market hours, likely also

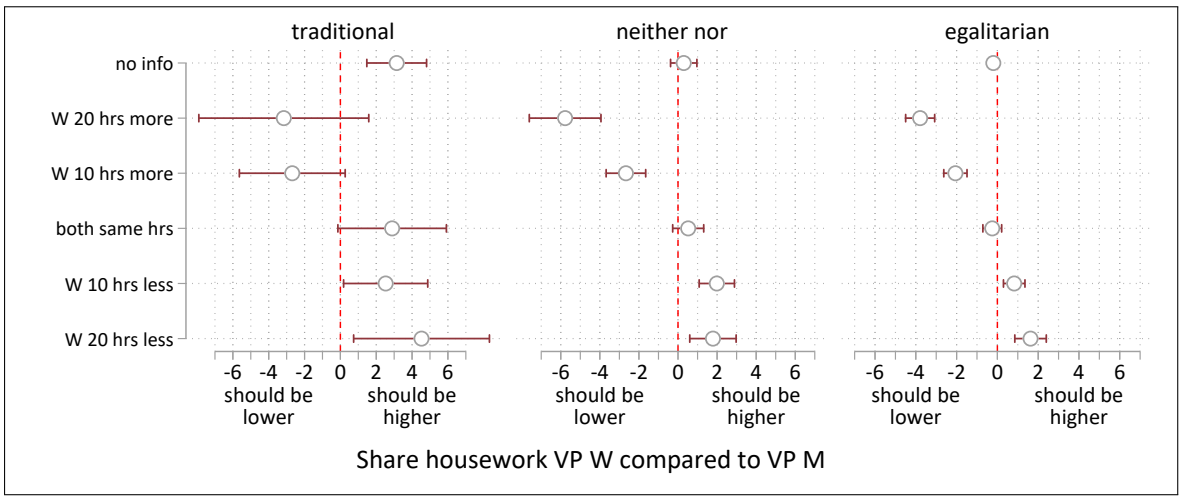


Figure 2.3. Effect of Adding Information on the Exact Relative Labor Market Hours by Respondent Group

Note: This figure shows the marginal effect of adding information on the vignette person's relative labor market hours on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column) and "egalitarian" (last column) respondents. Results were estimated separately by respondents' gender norms as measured by the item question "Men should participate in housework to the same extent as women." The corresponding regression tables can be found in Table 2.C2 in Appendix C.
Number of vignette evaluations of "traditional" respondents: 87; "neither nor" respondents: 439; "egalitarian" respondents: 1,907.

a consequence of the lower number of cases (only a relatively small group of respondents was classified as "traditional"). The effect of "no info" is on par with women having the same or a lower labor market participation (10 or 20 hours less than her partner), suggesting that in case explicit information is missing, women are expected to have at maximum the same, and frequently a lower labor market participation than their male partners. The effect is positive for all three levels (i.e., same, 10, or 20 labor market hours less), meaning women are expected to do more housework.[43] In the case of equal labor market hours, "traditional" respondents expect women to do more housework. This might indicate that respondents assume female vignette persons with equal labor market hours nevertheless make relatively lower contributions to the household income, which again reflects the standard case in Germany, where there is a pronounced gender pay gap. An additional analysis found support for this assumption (see Figure 2.A2 in Appendix A). In short, traditional respondents suggest women's share of housework should be higher when there is no information on her relative income. Again, this effect is on par with the effect for women earning 33% of the household income.

[43] In addition, "traditional" respondents expect women with a relatively higher participation in the labor market to do less housework than their male partner, which is clear evidence against the gender display thesis, which expects women who work more hours in the labor market or who have a higher income than their male partners to compensate for this deviance from the gender norm by doing most of the unpaid housework.

In sum, "egalitarian" and "traditional" respondents evaluations mostly differ in the case of low information: "Egalitarian" respondents seem to assume egalitarian labor market arrangements in which men and women have similar labor market hours and earnings, while "traditional" respondents tend to assume typical (German) gender inequalities, where women have less (or at least not more) labor market hours and lower income. In the case of full disclosure, all three groups appear to support equity norms.

These results are in line with the assumption that questions with low information content mainly measure different stereotypes, such as descriptive gender norms. However, they do not measure the intended prescriptive norms that women (men) should per se do more (less) of the housework. Given that respondents are provided with extensive information on the couples' labor market activities, all three types of respondents appear to support equity norms. If this interpretation is correct, the existing interpretation of evidence on the effect of gender norms on a couple's division on work has been misleading.

The analysis goes further to ensure any non-linear effects are not disregarded. Figure 2.4 shows the appropriateness of the housework share by gender of the vignette person. Technically speaking, a quadratic term of the vignette person's housework hours was included to account for a possible quadratic effect. In addition, this term was interacted with 1) her/his share of labor market hours (first row), 2) his/her share of income (second row), or 3) his/her share of child care (third row). The shares can either be smaller (first column), the same (second column), or larger (last column) than the respective partner's. As only vignettes of the high information condition were included, men and women on average show the same characteristics not only on the labor market, but also on all other dimensions.

Again, evaluations for vignette men and women hardly differ in most of the constellations and are stable for all. In the few constellations with a statistically significant difference by gender of the vignette person, it is always the male partner who is expected to do a larger share of housework, which is further evidence against traditional gender norms. Second, broad support is found again for equity norms: A housework share of roughly 50% is seen as appropriate in all constellations were both vignette persons have roughly the same shares of labor market hours, income, or child care (see second column of Figure 2.4). At the same time, vignette persons with a smaller (larger) share of working hours, income, or child care are expected to do more (less) of the housework (see first and last columns of Figure 2.4). Third, effects are symmetric for men and women: Both genders should decrease their time spent on housework when they have a larger share of working hours or income than their partners (see last column of Figure 2.4). This is clear evidence against the gender display thesis, and there are no signs of any non-linear effects that would contradict the evidence supporting the strong prevalence of equity norms. Finally, the last row of Figure 2.4 shows that respondents

do not only count labor market characteristics and housework hours against each other when rating the appropriate share of housework, but they also take the shares of child care into account. To our knowledge, this has not been tested before.

2.4.2 Robustness Checks

Replication with a Second Experiment
In a split sample, evaluations were collected on the appropriate share of the total workload (i.e., housework, paid employment, and child care if applicable) to test the robustness of the main results on (gender) norms. Is there evidence for normative beliefs that women should do a longer shift, and is this still (or less) true when they have equal labor market characteristics? The set-up of the experimental split on total workload was comparable as for housework with one main difference: There is no "low information" in this additional experiment, meaning that respondents always receive information about the labor market hours of both partners.

Overall, the results are comparable to the main results presented. Figure 2.A4 in Appendix A supports the finding that when information on the relative income is added, this shifts the evaluations further to the left, meaning that women should do less of the total workload, especially so for "traditional" respondents. The main findings are also supported by Figure 2.A5 in Appendix A. To control for non-linear effects, the appropriateness of the total workload was estimated by including interaction terms (cf. Figure 2.4). Throughout the graphs, it is clear that doing more of the housework allows both men and women to decrease their total workload. There are level differences by the amounts of labor market hours, relative income, and child care; however, there is no evidence for non-linear effects, especially not in cases were women are responsible for larger shares of couples' labor market hours or income. This is also evidence against the gender display thesis, which would have expected a U-shaped curve in situations in which women deviate from traditional gender norms.

Different Item-based Measurements of Gender Norms
In the main analyses, respondents were grouped into "traditional," "egalitarian," and "neither nor" attitudes. To do this, the item question closest to the experimental design was used, namely: "Men should participate in housework to the same extent as women." Based on this item question, only 3% of respondents were classified as "traditional." To ensure that the results were not biased by the classification itself, two additional classifications were made.

First, respondents were classified according to their agreement with a second item question: "Women should be more concerned about their family than about their career." Based on this item, 16% of respondents were classified as "traditional" and 41% as "neither nor". Results were similar to the main analysis: "Traditional" respondents

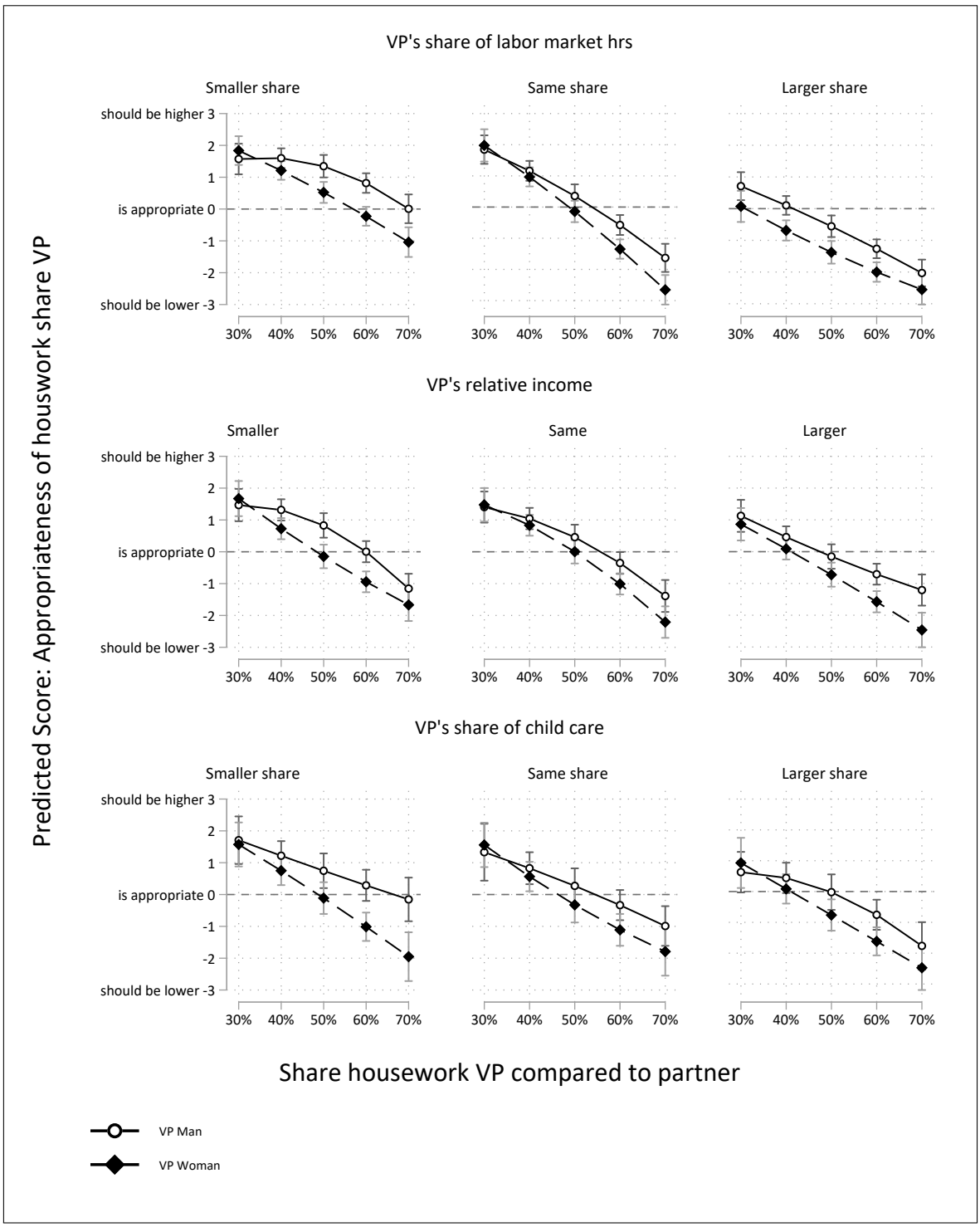


Figure 2.4. Appropriateness of the VP's Housework Share by Gender of the Vignette Person

Note: This figure shows the predicted appropriateness score of the vignette person's (VP's) housework share by his/her housework share under different conditions. Each graph shows the effect of the VP's housework share for a different amount of the VP's labor market hours (first row), his/her share of income (second row), and his/her share of child care (third row). Estimates were calculated based on separate regression models for labor market hours, relative income, and child care (see Table 2.C3 in Appendix C). To display separate predictions for male and female VPs, each regression model included interactions with the VP's gender.
Number of vignette evaluations for labor market hours and for relative income: 1,260; for share of child care: 841.

expect women to do more of the housework when no information on labor market hours and/or relative income is presented. Adding information again shifts the responses to the left (see Figure 2.A6 in Appendix A). Figure 2.A7 in Appendix A shows that broad support for equity norms for all three respondents groups was also found here once information on the labor market characteristics was presented. Women who work more hours in the labor market than their partners were expected to decrease their share of housework, which supports the rejection of the gender display thesis and the finding that equity norms trump gender norms.

Second, an index for gender norms was generated from three items from the gender role measure. With a Cronbach's alpha of only .56, the reliability of this index is relatively low, the main argument for not including it in the main analyses. Based on a tercile split, respondents were again divided into the three groups. Even with this broader definition of "traditional", conclusions are comparable to the main results (see Figures 2.A8 and 2.A9 in Appendix A).

Sample Restriction and Alternative Model Specifications

It is well known from survey research that respondents who complete questionnaires fastest are somewhat prone to method effects. For example, they might not take enough time to read the vignettes properly and therefore only remember the dimension they read last when answering related questions (e.g., Düval and Hinz, 2020). As this behavior would bias the results, the sample was restricted based on a proposal by Sauer et al. (2011)[44] to ensure the results are not biased based on the original sample restrictions. The results, which are not displayed here but are available on request, were again comparable to the main results. In a further check, all respondents not living in a relationship at the time of the interview were excluded, as they might have different perceptions of couples' division of paid and unpaid work. This did not change the results.

Finally, a model specification check was conducted by calculating all analyses using linear regression models to predict vignette evaluations.[45] As opposed to the main analysis, this check relied on ordinary least square regressions with cluster-robust standard errors to account for the hierarchical data structure. The results remained the same.

[44] The fastest 1% of respondents were excluded, along with all respondents who were two standard deviations faster than the average response time for the vignettes. In total, this affected only 17 respondents.

[45] The regression tables are not displayed, but are available on request.

2.5 Summary

This paper examines the conditionality of gender norms, beginning with a discussion of the limitations of standard item measures of gender norms and followed by an exploration of various mechanisms behind why respondents agree with traditional items on gender norms. The use of causal mediation analysis allowed for a decomposition of the gender effect into two different pathways: the effect of gender norms and the effect of equity norms. In a multifactorial survey experiment, respondents of the large scale German family panel pairfam were asked to evaluate the appropriateness of the division of housework of couples in different scenarios. The main results are summarized in four points in the following.

First, strong evidence was found for equity-norms that prescribe a gender neutral exchange of (economic) resources and housework. The appropriate share of housework strongly depended on monetary contributions and time availability: The higher the input in terms earnings or paid hours, the lower the share of housework that men and women are expected to do. Second, these analyses found no support for the gender norm that women should do a larger share of the housework than their male partners in general. On the contrary: When artificially equalizing men's and women's labor market resources in the experiment, women were found to be entitled to an even lower share of the housework, with a mean difference of only four percentage points. Third, no evidence for gender norms was found, neither that respondents expect women to do a comparatively larger share of housework, nor that respondents support gender-specific exchange rates between economic resources and housework. Finally, the results show that only in low information conditions do ("traditional") individuals base their evaluations on stereotypic assumptions about men's and women's earnings contributions. After providing information on labor market hours and/or relative income, no effect of gender remains. Women are therefore only expected to do more of the housework when they provide less of the family's income or work less hours in the labor market. This is strong support for the conditionality of gender norms. Even more so, this analysis questions the existence of a direct effect of gender norms, simultaneously suggesting the standard measurement of gender norms to be artificial.

2.6 Conclusion

This analysis began with the awareness that gender norms are seen as one of the main predictors of gender inequalities insofar that they prescribe women the main responsibility for domestic tasks. However, these results clearly support the notion that norms of equity (meanwhile) dominate gender norms in Western societies. This was overlooked in prior research, most likely due to an overestimation of the prevalence

of gender norms by way of vague item questions and the entanglement of prescriptive gender norms with norms of equity. In this analysis, even respondents classified as "traditional" by the standard items no longer supported traditional arrangements once provided with extensive information on individuals' labor market status. In sum, these results suggest a pronounced conditionality of gender norms and questions the standard measurement thereof.

Researchers must be more specific when discussing gender norms as an explanation for gender inequalities (for a seminal study, see Thébaud et al., 2021). Based on these findings, it is unlikely that there is a gender norm that determines individuals' actual division of housework. This is consistent with the general literature on norms and attitudes, where there is much debate about whether norms and attitudes indeed have a strong influence on behavior, especially in high-cost situations (c.f. Ajzen, 1991). It is thus particularly important to describe the assumed mechanisms accurately so as to avoid confusing gender norms and other (non-gender) norms such as equity. A further take-away is that the future research would profit from a more precise discussion of what is meant and measured when speaking about gender norms or gender ideologies: Are these, for instance, descriptive or prescriptive norms?

In particular, researchers interested in cross-country or multi-group comparisons must be aware that differences in the support of gender norms (often referred to as "role attitudes") might not indicate variance in gender ideologies. Instead, they might reflect differences in gender inequalities as they actually exist. The practical recommendation to avoid these issues is to always use the same amount (and kind) of information for all respondents; ideally, to include more detailed information in single item questions as well.

This study is not without limitations. This particular design allowed for the study of the amount of daily housework chores male and female intimate partners are expected to do, but not gender norms regarding the specialization on "male" versus "female" tasks (Schneider, 2012). However, the vast majority of respondents supported the view that couples should decide on their work division themselves.[46] This might suggest that egalitarian views also hold for other aspects in the organization of housework.[47] The external validity of these results could also be limited, as only younger cohorts are included in the pairfam survey: the oldest age cohort was born 1971 (i.e., the maximum age at time of the survey was 47), and the mean age of respondents in wave 10 was 35. This homogeneous age sample restricted the variance in gender norms. Further studies

[46] In the vignette task, this opinion was reflected in the large number of respondents that evaluated the indicated work division as appropriate; in addition, this opinion was measured with a direct item question.

[47] This is supported by an additional experiment in which respondents were asked to evaluate the total work division (instead of the housework share). No gender effects were found (see section 2.4.2).

might research whether these results can be generalized to older respondents, and also whether the age-attitudes nexus found in prior research (Mays, 2012) still holds when using more informative survey questions.

Another direction for future research is to repeat this experiment in other national contexts. It is important to understand whether traditional item questions are biased in countries that show a higher labor market participation of women or mothers (such as the U.S.). Cross-country research with more informative questions on gender norms would not only present an important methodological contribution, but would also provide more firm conclusions regarding the sources of variable gender inequalities (and how to reduce them with policy measures).

Finally, the validity of these results could also be threatened by a social desirability bias. Methodological research suggests that respondents in factorial surveys display fewer socially desirable reactions compared to responses to single item questions (Auspurg and Hinz, 2015). This experiment only implemented a between-respondent variation of the gender of the vignette person. In addition, the vignettes were administered in a self-completion survey mode. Both features are known to minimize social desirability bias (Groves et al., 2004). More importantly, one can argue that it is not at all plausible that a social desirability bias could have distorted the measurement of widely shared norms. Normative beliefs with such a low desirability that they cannot even be expressed in an anonymous interview situation certainly do not represent the common view. This boosts confidence that the results in fact capture common norms.

What else, if not gender norms, could explain the gender inequalities in housework? It is worth mentioning that in this experiment, men's and women's labor market resources were artificially equalized. In the real world, (unobserved) differences in endowments might represent partial explanations. In addition, it is important to recall that gender is a social construct that is not restricted to gender norms. Doing (or avoiding) housework could simply help individuals to signal their gender identities (Bittman et al., 2003; Thébaud et al., 2021). These self-definitions are, however, conceptually different from whether people approve of traditional or egalitarian gender relations (Wood and Eagly, 2015). In particular in a 'liberal' context with high levels of gender equality in the political, economic, and educational spheres, individuals might make use of gender stereotypes in other realms to affirm their gender identities (see also the findings on the "paradox in gender equality"; e.g., Stoet and Geary, 2018).

Following these results, equity norms clearly trump gender norms. However, this does not necessarily imply that gender as a social construct no longer primes inequalities (in housework shares). Survey experiments such as the one described in this study could better explore other related concepts such as gender identities (Akerlof and Kranton, 2000; Kroska, 2000) or any other mechanisms underlying men's and women's distinct (housework) practices.

Appendices

In Appendix A, we provide some additional analyses, e.g., descriptive results, results that support our main conclusion but were not included into the main article, and our robustness checks. Appendix B provides more information on existing experimental research on gender inequality and additional information our experimental set-up. And in Appendix C you find the regression tables.

Appendix A: Additional Results

Table 2.A1. Descriptive Statistics

	N	Min	Max	Mean	s.d.
Resp. gender (1: woman)	1,232	0	1	0.550	0.498
Resp. age	1,232	24	47	0.350	8.430
Resp. relationship (1: yes)	1,228	0	1	0.730	0.444
Item question	1,232	0	1	4.260	0.871
Vignettes: Total	3,693	-5	5	0.040	2.100
Vignettes: Female VP	1,835	-5	5	-0.125	2.080
Vignettes: Male VP	1,858	-5	5	0.203	2.100

Descriptive Analyses

Table 2.A1 shows descriptive statistics on our sample. About 55% of the sample is female. The average age of the respondents is 35 years, and around 73% or respondents are in a relationship. The left side of Figure 2.A1 shows the distribution of the standard item question on gender norms our analyses mainly focus on: "Men should participate in housework to the same extent as women." Respondents could answer on a 5-point scale ranging from 1 "Disagree completely" to 5 "Agree completely". Respondents that answered 1 or 2 are classified as traditional, those that answered 4 or 5 are egalitarian, while the others (answer 3) are neither nor. With the mean being 4.3 (*s.d.*: 0.9) on the 5-point rating scale, most respondents agreed with the statement that men should participate to the same extent in housework as women (see Table 2.A1). Nearly 80% of respondents were classified as showing "egalitarian" gender norms. Only 3.2% classify as strictly "traditional," while the remaining 17.5% are "neither nor."

The right side of Figure 2.A1 shows the distribution of the responses to the multifactorial survey experiment. Remember, on average male and female vignette persons do the same amounts of housework (mean share of 50%) and show exactly the same economic resources (labor market hours and relative income). Therefore, a mean rating of zero would indicate, that male and female vignette persons should do the same amounts of housework. With a mean evaluation of 0.20 (*s.d.*: 2.1) for male vignette

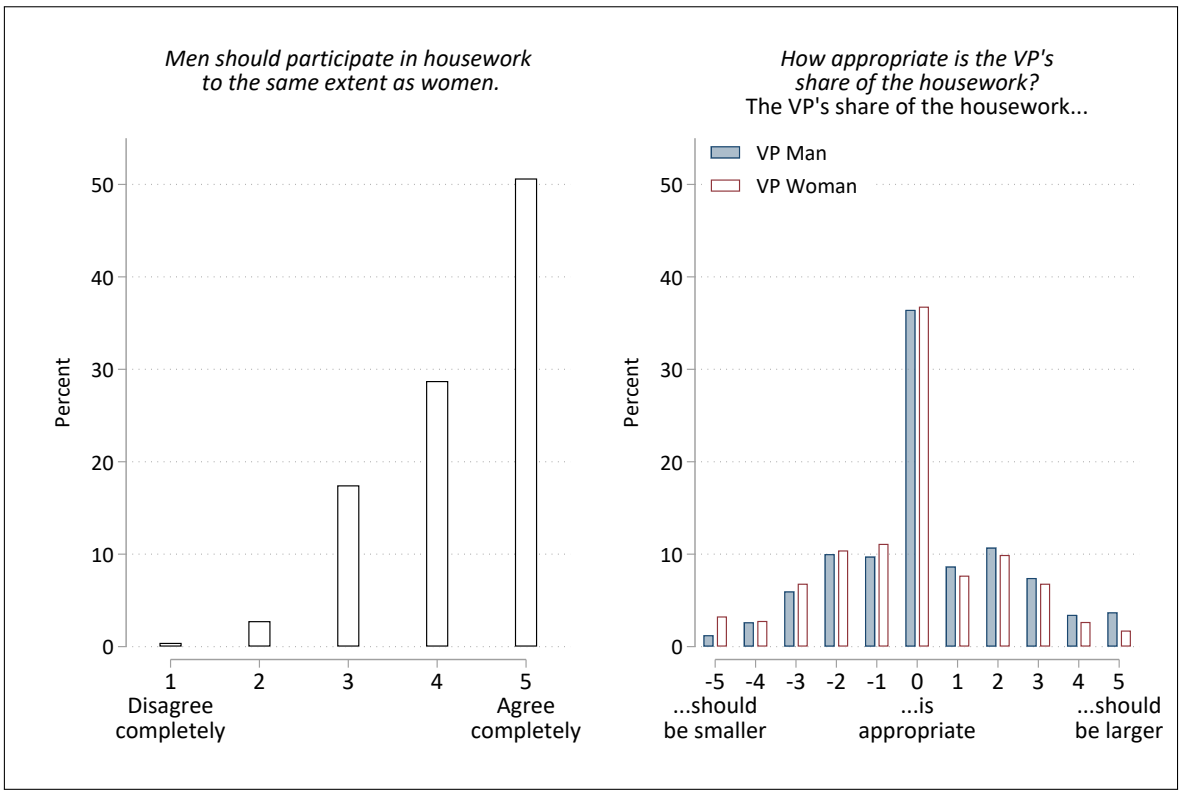


Figure 2.A1. Distribution of Responses: Item Question and Factorial Survey Experiment

Note: This figure shows distribution of the responses on the item question on gender norms (left) and on the multifactorial survey experiment (right).
Number of respondents for the item question: 1,232; number of vignette evaluations: 3,693, thereof 1,835 for female and 1,858 for male vignette persons.

persons and of -0.13 (*s.d.*: 2.1) for female vignette persons, we hardly see any difference by gender. This is first indication for a strong support of equity. On average, women should even do 4 percentage points less of the housework than men. But keep in mind, this is the average rating over all vignette dimensions and information conditions.

Additional Analyses

To support our assumption that respondents assume female vignette persons with equal labor market hours nevertheless to make relatively lower contributions to the household income, we run additional analyses. The results are shown in Figure 2.A2. Note that for the cases shown here, also full information on the labor market hours was provided (vignettes with information on income always included information on the labor market hours, see section 2.3.1). Here, in the full information condition, also "traditional" respondents no longer show a statistically significant gender effect (i.e., also these respondents rated the appropriate share of housework to be the same for male and female vignette persons) in case the vignettes sketched both partners as having the same labor market resources (same income; and, at least on average, also same

labor market hours). Furthermore, and even more interesting, we can conclude form the results in Figure 2.A2 that "traditional" respondents in the no information condition again support "traditional" arrangements where the female partner makes a lower contribution to the household income.[48] Note also, that only this traditional income arrangement makes the "traditional" respondents supporting a higher housework share for the female vignette person. However, even though the effect size is quite large, this marginal effect is not statistically significant anymore ($p = 0.058$). This is probably due to the relatively small number of "traditional" respondents.

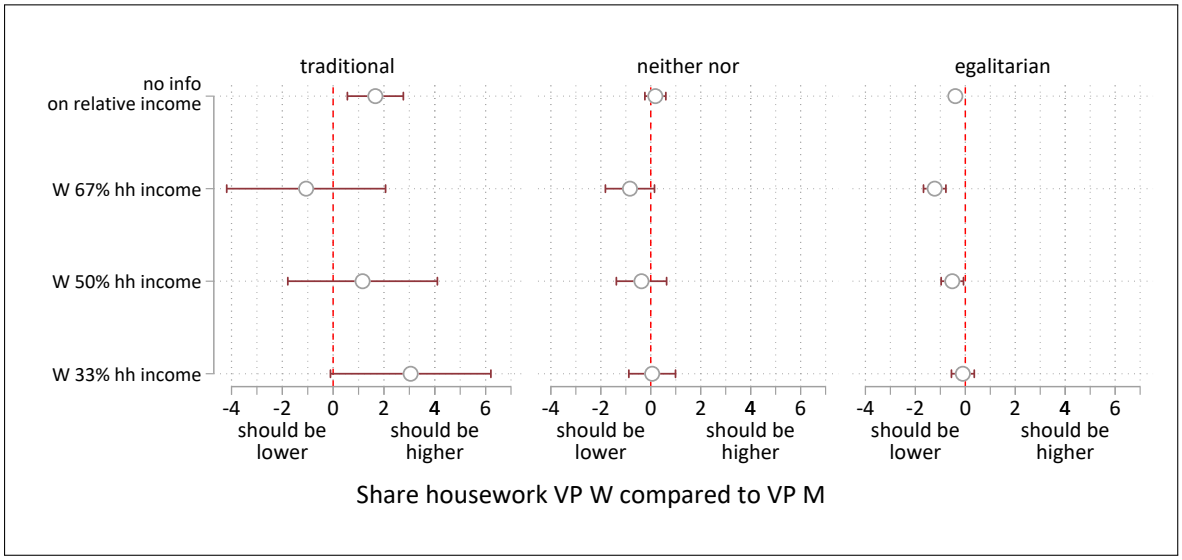


Figure 2.A2. Effect of Adding Information on the Exact Relative Income by Respondent Group

Note: This figure shows the marginal effects of adding information on the vignette person's relative income on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column) and "egalitarian" (last column) respondents. The results were estimated separately by the respondents' gender norms measured by the item question "Men should participate in housework to the same extent as women." The corresponding regression tables can be seen in Table 2.C4 in Appendix C.
Number of vignette evaluations of "traditional" respondents: 117; of "neither nor" respondents: 647; of "egalitarian" respondents: 2,929.

For "egalitarian" respondents, we can again conclude that these respondents assumed egalitarian arrangements; the effects found for the no information condition is closest to the effect found for vignettes where male and female vignette persons were described as making an egalitarian contribution to the household income. As can be expected, the "neither nor" group is in between egalitarian and traditional respondents. For them, the "no info" condition is mostly on par with the condition where the woman earns half as much as the man. However, none of the effects is statistically significant. Over all respondents groups, the effects of income are in line with equity norms; but effect sizes

[48] The effect for the no information condition is mostly on par with the condition where the female vignette person was sketched as making a lower monetary contribution.

are much lower than those found for labor market hours; this can be seen as evidence for time availability instead of bargaining theories.

Further evidence for our main results can be found in Figure 2.A3. Here, we regressed all vignette dimensions on the appropriateness of the vignette person's share of housework. Again, our results are in line with equity norms: Doing more labor market hours or having a higher relative income allows for a lesser share of the housework. At the same time, being responsible for lesser shares on the labor market results in having to do a larger share of the housework. Furthermore, we find no statistically significant difference between male and female vignette persons.[49] This means that respondents do not differ in their evaluations for men and women. This is further evidence against gender norms.

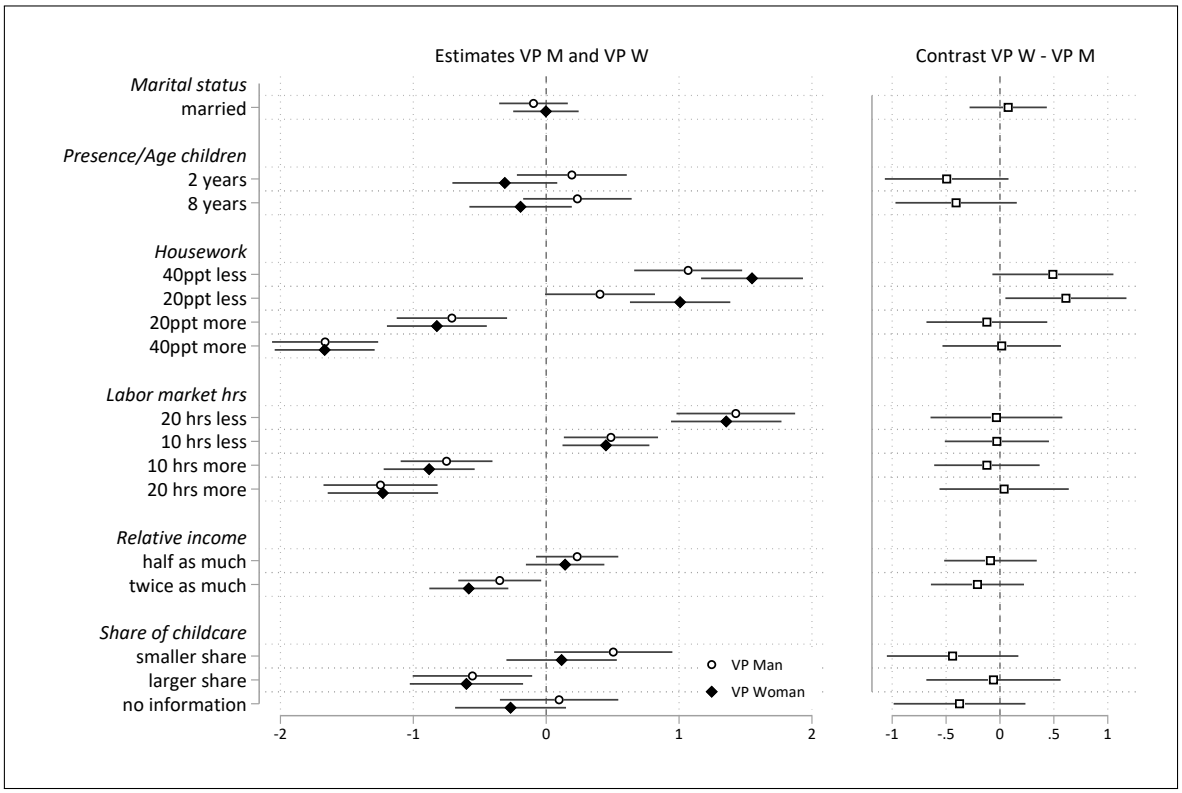


Figure 2.A3. Impact of Vignette Dimensions by Gender of Vignette Person

Note: This this figure shows regression coefficients for all vignette dimensions on the appropriateness of the vignette person's share of housework. The coefficients are displayed separately by gender of the vignette person. The left panel regressions were estimated separately for male and female vignette persons. The right panel shows the difference in effect sizes across both genders. This figures includes only vignettes of the high information condition. The corresponding regression tables can be seen in Table 2.C5 in Appendix C.
Number of vignette evaluations: 3,693, thereof 1,835 of female and 1,858 of male vignette persons.

[49] We only find one statistically significant difference (see right side of the graph, $p = .033$), which is nearly in the range of what can be expected to be statistically significant by chance.

Robustness Checks

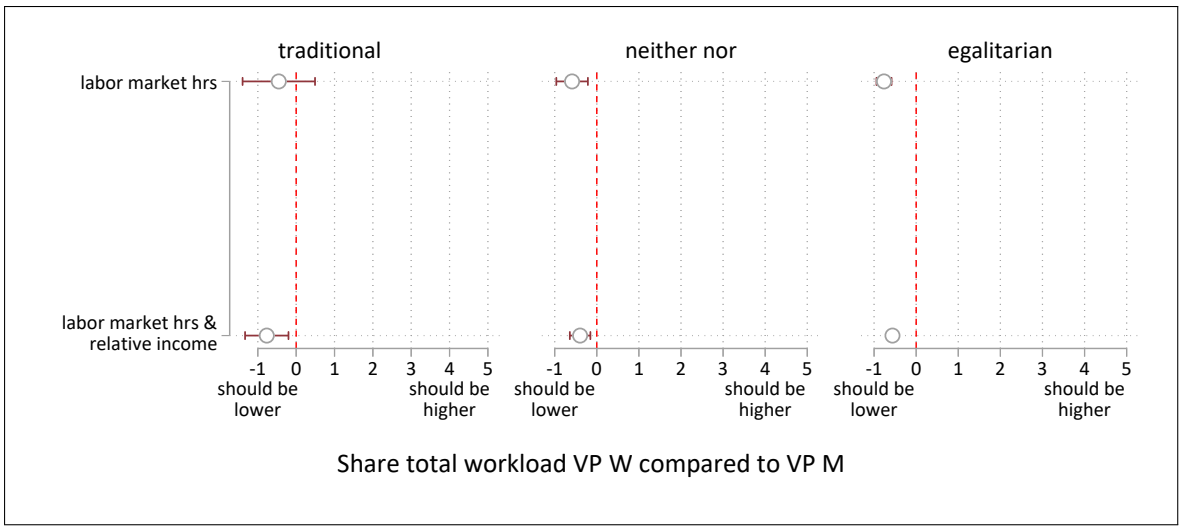


Figure 2.A4. Effect of Adding Information on Labor Market Characteristics by Respondent Group (Experiment on Total Workload)

Note: This figure shows the marginal effect of adding information on the vignette person's relative income on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column) and "egalitarian" (last column) respondents. This figure is based on an additional factorial survey experiment on the vignette person's total workload. The regression results are not displayed, but are available on request.
Number of vignette evaluations of "traditional" respondents: 374; of "neither nor" respondents: 1,624; of "egalitarian" respondents: 7,862.

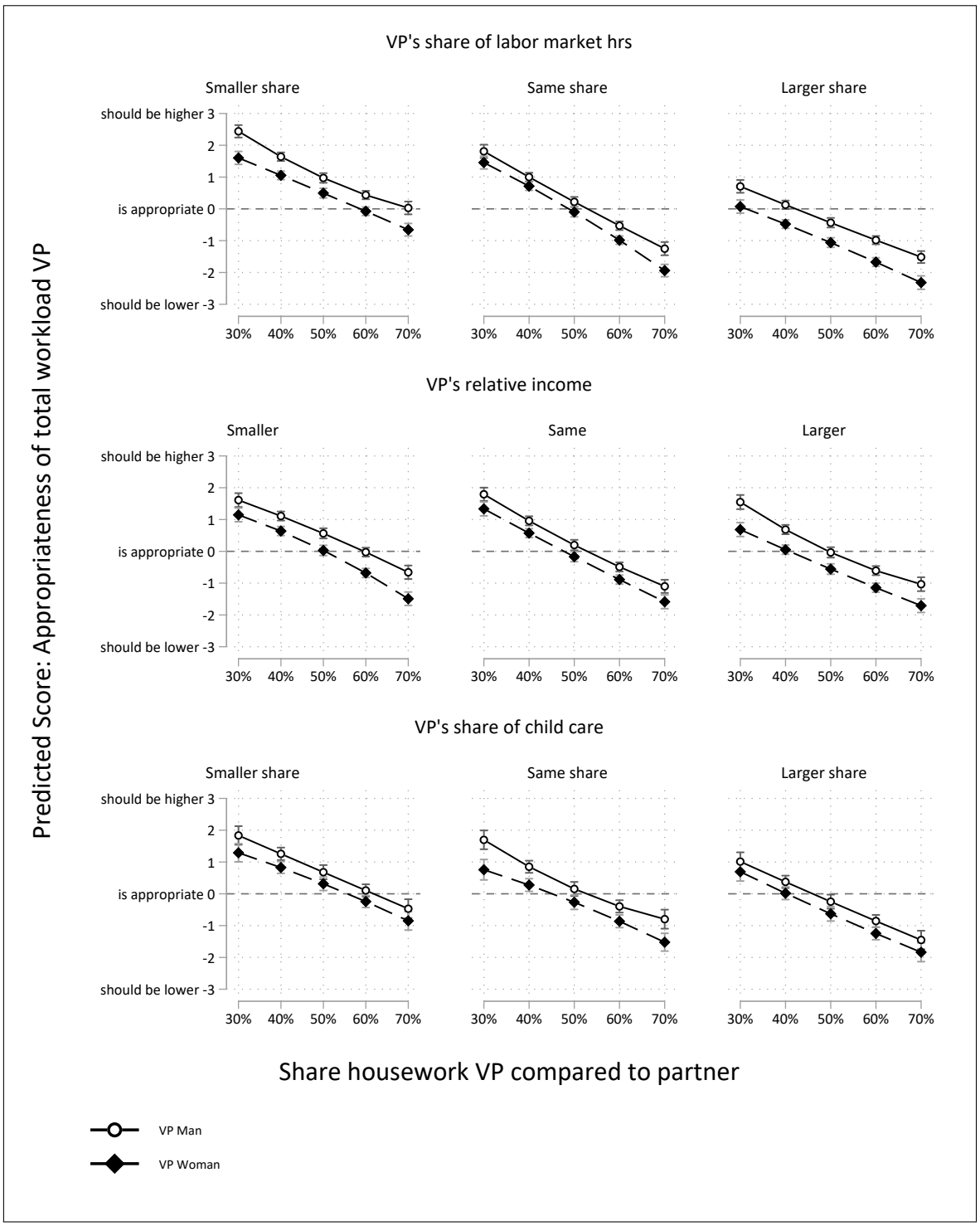


Figure 2.A5. Appropriateness of the VP's Housework Share by Gender of the Vignette Person (Experiment on Total Workload)

Note: This figure shows the predicted score of the appropriateness of the vignette person's (VP's) total workload by his/her housework share under different conditions. Each graph shows the effect of the VP's housework share for a different amount of the VP's labor market hours (first row), his/her share of income (second row), and his/her share of child care (third row). The predictions were calculated based on separate regressions for labor market hours, relative income, and child care. To display separate predictions for male and female vignette persons, each regression included interactions with the gender of the vignette person. The regression results are not displayed, but are available on request.
Number of vignette evaluations for labor market hours and for relative income: 6,858; for share of child care: 4,548.

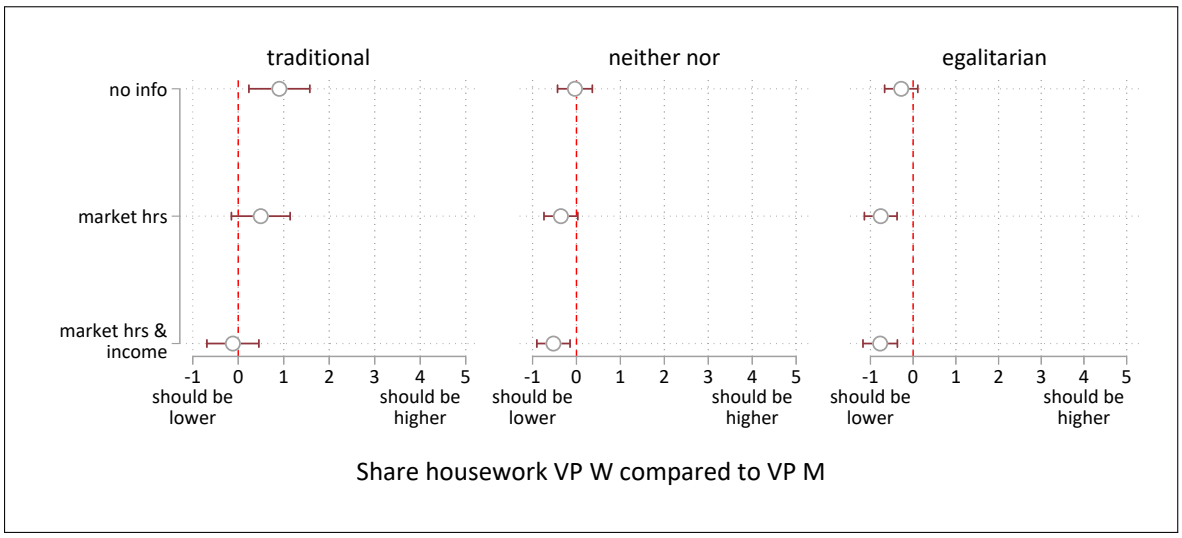


Figure 2.A6. Effect of Adding Information on Labor Market Characteristics by Respondent Group (Classification by Item Question 2)

Note: This figure shows the marginal effect of adding information on the labor market characteristics ("no info," "labor market hours," and "labor market hours and relative income") on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column), and "egalitarian" (last column) respondents. For this figure, respondents are separated into the three gender norm groups by using the item question "Women should be more concerned about their family than about their career." The regression results are not displayed, but are available on request.
Number of vignette evaluations of "traditional" respondents: 588; of "neither nor" respondents: 1,527; of "egalitarian" respondents: 1,572.

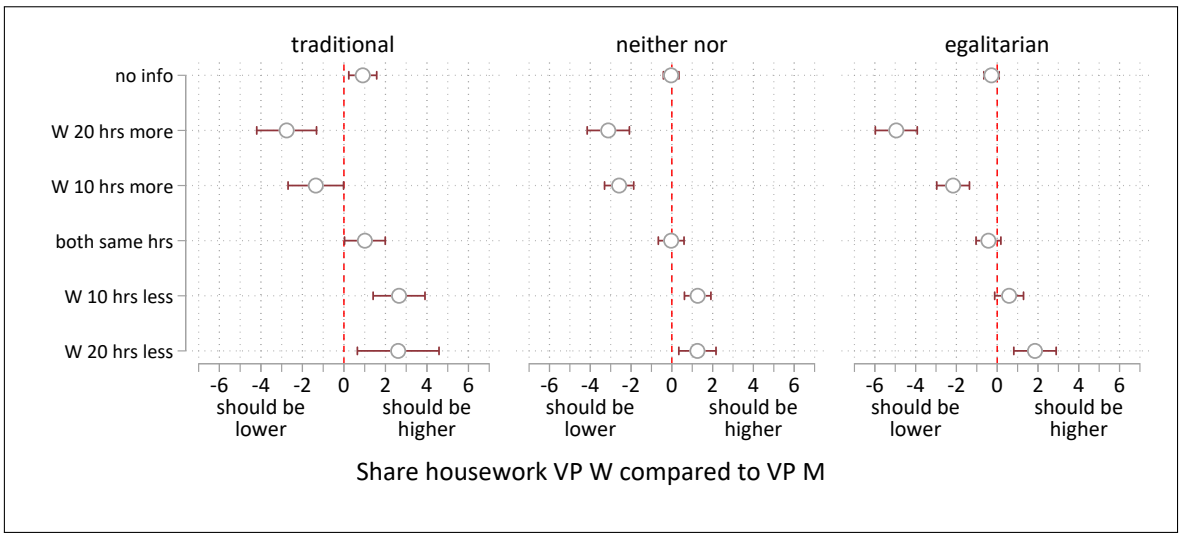


Figure 2.A7. Effect of Adding Information on the Exact Relative Labor Market Hours by Respondent Group (Classification by Item Question 2)

Note: This figure shows the marginal effect of adding information on the vignette person's relative labor market hours on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column) and "egalitarian" (last column) respondents. For this figure, respondents are separated into the three gender norm groups by using the item question "Women should be more concerned about their family than about their career." The regression results are not displayed, but are available on request.
Number of vignette evaluations of "traditional" respondents: 588; of "neither nor" respondents: 1,527; of "egalitarian" respondents: 1,572.

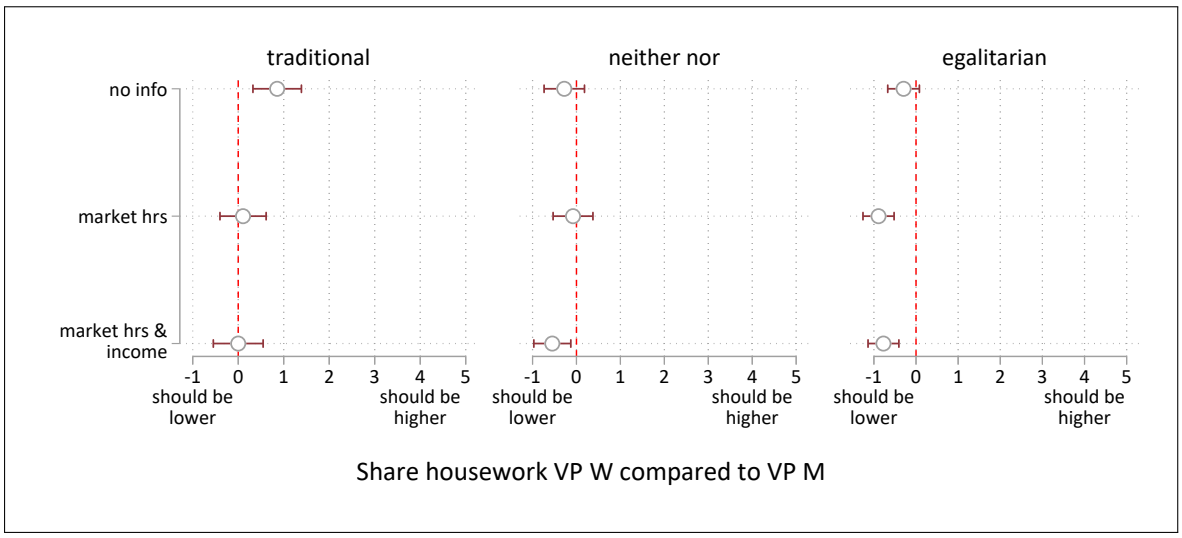


Figure 2.A8. Effect of Adding Information on Labor Market Characteristics by Respondent Group (Classification by Index on Gender Norms)

Note: This figure shows the marginal effect of adding information on the labor market characteristics ("no info," "labor market hours," and "labor market hours and relative income") on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column), and "egalitarian" (last column) respondents. For this figure, respondents are separated into the three gender norm groups by using an index based on three item questions on gender norms. The regression results are not displayed, but are available on request.
Number of vignette evaluations of "traditional" respondents: 861; of "neither nor" respondents: 1,142; of "egalitarian" respondents: 1,684.

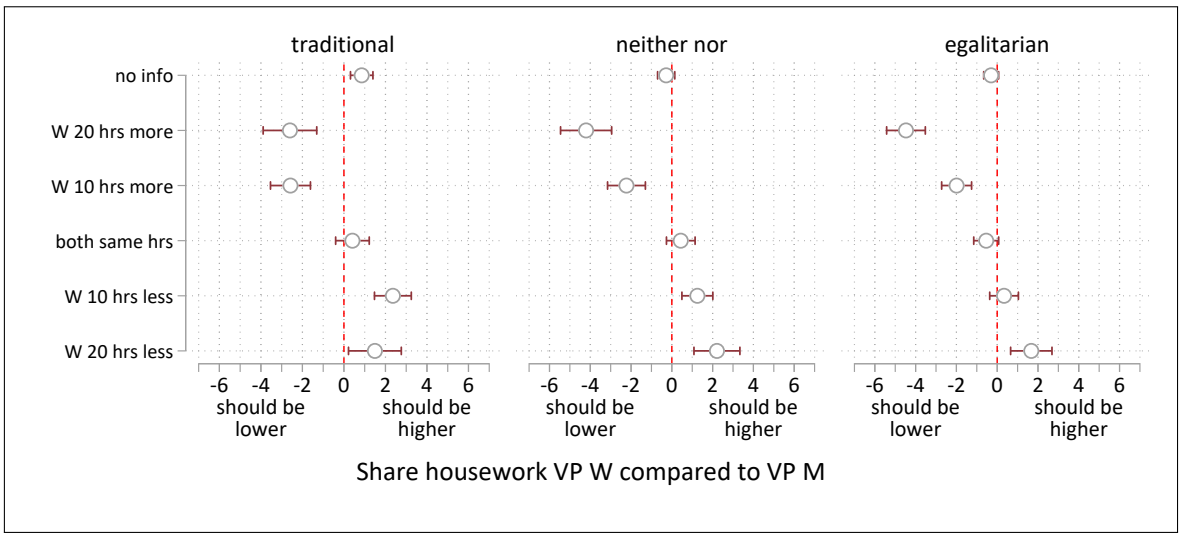


Figure 2.A9. Effect of Adding Information on the Exact Relative Labor Market Hours by Respondent Group (Classification by Index on Gender Norms)

Note: This figure shows the marginal effect of adding information on the vignette person's relative labor market hours on the appropriateness of the vignette women's share of housework compared to the vignette men's. The effects are shown separately for "traditional" (first column), "neither nor" (second column) and "egalitarian" (last column) respondents. For this figure, respondents are separated into the three gender norm groups by using an index based on three item questions on gender norms. The regression results are not displayed, but are available on request.
Number of vignette evaluations of "traditional" respondents: 861; of "neither nor" respondents: 1,142; of "egalitarian" respondents: 1,684.

Appendix B: Previous Research and Experimental Design

Further Information on Existing Research

In section 2.2.3, we provided a short overview of existing experimental research on preferences on ideal work divisions (Auspurg et al., 2017b; Pedulla and Thébaud, 2015) gendered descriptive norms (Auspurg et al., 2017a; Correll et al., 2007), and different housework standards and/or fairness norms men and women are judged by (Carriero and Todesco, 2017; Schulz, 2021; Thébaud et al., 2021). In this Appendix, we provide more information on this existing research.

First, Pedulla and Thébaud (2015) and Auspurg et al. (2017b) focused on preferences for more or less gendered work arrangements. Interestingly, these studies that were based on experimental instead of observational data found that individuals have a strong favor for gender-blind, egalitarian arrangements. For the U.S., Pedulla and Thébaud (2015) found that irrespective of their gender, respondents prefer an egalitarian division of paid and unpaid work. Auspurg et al. (2017b) used a factorial survey experiment on a British sample to separate gender from other explanations for a preferred division of work. Their main result is that it is not gender that influences preferences for work divisions. Instead, respondents base their preferences on equitable distribution rules. These preferences were, however, bound to options that easily allow for these options, e.g. by removing (institutional) constraints such as a lack of child care facilities or earnings inequalities that might hamper dual-earner arrangements in reality. Jacobs and Gerson (2016) assigned respondents in the U.S. (Time Sharing Experiments in the Social Sciences, TESS) to three different scenarios (vignettes) with varying family constellations, and asked them whether they think the described single or married mothers or fathers should continue to work full time, stay at home, or scale back to part time. The support of both mother's and father's employment was found to substantially increase when the persons were described as being satisfied with their job, or when respondents were made to believe that the family depended on their income. There was also a moderate gender gap in the expected direction (respondents supported the employment of fathers overall to a larger extent than the employment of mothers), but this effect was small in comparison to the effects of the economic variables.[50]

Second, experimental studies on (just) earning differences between men and women (Auspurg et al., 2017a) and mothers and childless women (Correll et al., 2007) found support for a gendered status value or gendered descriptive norms. Women and/or mothers are perceived to be less competent and commited to work in the labor market (Correll et al., 2007). Auspurg et al. (2017a) found that experienced inequalities

[50] The experiment was also not explicitly designed to test gender norms (in the vignettes, there was, e.g., only vague information on the economic situation). The remaining gender gap might therefore simply reflect respondents beliefs about differences in male's and female's salaries.

in men's and women's earnings influences justice perceptions and ratings. Gendered status beliefs are internalized by men and women and, hence, considered fair (Auspurg et al., 2017a). Since the results only refer to descriptive beliefs about gender differences in income or in the labor market, these results are of course not transferable to prescriptive norms about the division of labor in the household. However, they do shed light on the fact that gender-specific status beliefs or descriptive gender norms have an impact on what is considered fair or appropriate.

Third, Carriero and Todesco (2017) and Schulz (2021) used factorial survey experiments to probe on fairness norms men and women use to judge couples' division of paid and unpaid work. In different regions in North-West Italy, Carriero and Todesco (2017) probed on the perceived fairness of different housework constellations of hypothetical couples. The fair amount of housework was found to depend on the working hours and gender—but in the opposite direction than expected by traditional gender norms. Also according to the authors, this result likely represented an artefact caused by a too low gradation regarding the share of housework, where the respondents could only choose between three extreme categories (female does 20%, 50%, or 80% of the housework). In a sample of German residents, Schulz (2021) compared men's and women's attitudes toward division of housework. He found that labor market characteristics influenced of what men and women perceive as fair. As he found that men should do more of non-routine housework tasks like repairs or paper work, also gender seems to matter. Schulz (2021) did not vary the amount of information presented to the respondents. Therefore, the conditionality of gender norms cannot be tested with his design. He can only measure if there is a direct effect of gender after controlling for labor market characteristics. However, as relevant factors (e.g., information on the couples' division of child care) were not included in his experimental set-up it is unclear if a (remaining) gender differences points toward the existence of a gender norm or if respondents simply assumed, e.g., a larger share of child care for women, and, hence, think it is fair if she has to do less of the housework.

Finally, in a seminal study, Thébaud et al. (2021) provided insights on different housework standards men and women are held accountable for. They used a factorial survey experiment in which respondents from the U.S are presented with pictures of clean and messy rooms of male and female occupants. The authors found an effect of the gender of the room occupant insofar that women who occupy a tidy room are judged as less clean and less moral than men who occupy a tidy room. In messy rooms, no gender difference was found. They explain these results by women facing negative stereotypes irrespective of the situation, whereas for men only messiness activates negative stereotypes. Especially interesting in context to our study is that Thébaud et al. (2021) found gender differences in expected responsibility for housework. If they work full time, women should do more of the housework. This is true both in situations

where they live alone or when they are a parent and live with a spouse. However, the authors did not experimentally manipulate the work and family conditions when they asked about the housework responsibilities. This is where our study comes in. In our factorial survey experiment we do not only experimentally vary the family conditions (e.g., marital status and presence and age of children), in addition we vary a wide amount of work conditions (e.g., both partner's shares of housework, child care, and hours in the labor market and relative income). This allows us to probe if (women's) expected responsibility for housework depends on family conditions, work conditions or interactions of these spheres.

Information on the Experimental Design

In our multifactorial survey experiment, we asked respondents to evaluate couples that were described by seven dimensions that are all known to influence couples' decision on work shares (see, e.g., Baxter et al., 2008; Davis and Greenstein, 2009): (1) the couple's marital status; (2) the presence and age of children in the household; (3) the distribution of child care between partners; (4) the share of housework done by both partners; (5) the male partner's labor market hours; (6) the female partner's labor market hours; and (7) the relative income of both partners. Between two and five categories (levels) were defined for each of the seven dimensions (see Table 2.B1 for all dimensions and levels).

Additionally, we varied the amount of information that was presented to the respondents: (A) in the low information condition, we only presented information on family status, child care, and the couple's housework share (i.e., dimensions 1-4). The remaining dimensions were blanked. This question format closely resembles the one used in the classical item questions. (B) In the medium information condition, we additionally presented information on both partners' labor market hours (i.e., dimensions 5 and 6), and (C) in the high information condition we presented all seven dimensions, which also includes information on the relative income of both partners (i.e., dimension 7). See Figure 2.B1 for an example vignettes of the three information conditions.

This between-respondent split allowed us to probe if respondents only agree with the traditional housework arrangement if they miss information on couple's labor market resources like their labor market hours and relative income. This would hint toward the conditionality of gender norms. In a second between-respondent split, we implemented a variation of the evaluation task: About half of the respondents throughout had to evaluate the housework share of a female vignette person, while the other half evaluated the housework share of a male vignette person. We deliberately used a between-respondent split to avoid social desirability bias, and also to avoid confusion when respondents had to switch between the evaluation of men's and women's chores. Respondents were randomly allocated to both between-respondent splits (i.e.,

Table 2.B1. Vignette Dimensions and Levels

	Dimensions	Levels	Information condition
1	Marital status	Unmarried/ married	Low
2	Presence and age of child	No children/ 2 years/ 8 years	Low
3	Share of child care (relative to partner)	Larger share/ Smaller share/ Same share/ No info	Low
4	Share of housework per week	70% (21h)/ 60% (18h)/ 50% (15h)/ 40% (12h)/ 30% (9h)	Low
5	Labor market hours per week (man)	40 h/ 30 h/ 20 h/ No info	Medium
6	Labor market hours per week (woman)	40 h/ 30 h/ 20 h/ No info	Medium
7	Relative income (relative to partner)	Twice as much/ Half as much/ Same/ No info	High

information condition and evaluation task).

The full set of all possible scenarios (the vignette universe) spans all 7,680 possible combinations of the dimension levels. To select our fraction of 750 vignettes, we used a *D*-efficient sampling technique that minimizes correlations between dimensions while maximizing the variance of each of the dimensions (for details: Atzmüller and Steiner, 2010; Auspurg and Hinz, 2015). This ensured that all main effects of the vignette dimensions and all two- and three-way interactions between dimensions were not confounded with each other. Put differently, there exists no gender bias in our vignette sample, hence, in our fictional vignette world gender does not influence the labor market characteristics of men and women. In the realized sample, the correlation of gender with income was $r = .001$ ($p = .970$) and with labor market hours was $r = -.032$ ($p = .107$).

We split the sample of 750 vignettes into 250 different questionnaire versions, each containing 3 vignettes. The different questionnaire versions were allocated randomly to the respondents, with the ordering of the three scenarios being randomized for each respondent to neutralize possible effects of the vignette order (Auspurg and Jäckle, 2017). Our experiment was administered in a self-completion mode via computer-assisted self-interviews (CASI). Self-completion is the recommended mode for factorial survey experiments of this type, first because the vignettes might be better understood by respondents if they read them directly than if they are read out by an interviewer, and second, because this mode reduces possible social desirability bias (Auspurg and Hinz, 2015). A number of pretests conducted prior to the implementation in pairfam suggested that respondents coped well with the questions and the level of complexity. For more information on the vignette module, see Düval and Auspurg (2020).

Split 1: A married couple has an 8-year-old child.

Both are normally responsible for 50% (15 hours a week) of the weekly housework (e.g., laundry, cooking, cleaning, repairs).
She is responsible for a smaller share of child care than is he.

Split 2: + She works 30 hours per week, he works 40 hours per week.

Split 3: + Her contribution to their monthly household income is approximately half of his.

How appropriate do you think her share of the housework is?

Her share of the housework…

should be much smaller					is appropriate					should be much higher
-5	-4	-3	-2	-1	0	1	2	3	4	5
o	o	o	o	o	o	o	o	o	o	o

Figure 2.B1. Example Vignette with Information on All Three Information Splits

Note: pairfam Group (2019). For the wording of the original German vignettes, see Düval and Auspurg (2020).

Appendix C: Regression Tables

Table 2.C1. RI-Regression Model of the Appropriateness Rating: The Effect of Adding Information on Labor Market Characteristics by Respondent Group

	(1) Traditional resp.	(2) Neither nor resp.	(3) Egalitarian resp.
Main effects: vignette factors			
VP woman (ref: VP man)	3.142***	0.295	−0.199
	(4.009)	(0.853)	(−1.400)
Information condition (ref: low info)			
Labor market hrs	2.081**	0.302	0.205
	(2.715)	(1.000)	(1.417)
Labor market hrs + relative income	1.544	0.185	0.0326
	(1.856)	(0.591)	(0.234)
Interactions: VP woman X			
Labor market hrs + relative income	−2.086	−0.649	−0.421*
	(−1.762)	(−1.389)	(−2.117)
Labor market hrs	−2.385*	−0.198	−0.409*
	(−2.280)	(−0.449)	(−2.004)
Constant	−2.267***	−0.115	0.207*
	(−3.687)	(−0.497)	(2.116)
Number of observations	117	647	2,929
Number of individuals	39	221	984
Wald test χ^2	19.76	4.71	43.47
p-value	0.001	0.453	0.000
ρ	0.055	0.151	0.065

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(5)$ for all three columns.

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.C2. RI-Regression Model of the Appropriateness Rating: The Effect of Adding Information on Exact Labor Market Characteristics by Respondent Group

	(1) Traditional resp.	(2) Neither nor resp.	(3) Egalitarian resp.
Main effects: vignette factors			
VP woman (ref: VP man)	3.142***	0.293	−0.199
	(3.694)	(0.852)	(−1.465)
Labor market hrs (ref: no info)			
W 20 hrs more	3.496**	3.439***	2.017***
	(2.873)	(4.823)	(7.321)
W 10 hrs more	4.202***	1.786***	1.035***
	(3.452)	(4.059)	(4.441)
Both same hrs	2.326*	0.023	0.024
	(2.117)	(0.063)	(0.120)
W 10 hrs less	−0.002	−0.429	−0.698**
	(−0.002)	(−1.108)	(−3.160)
W 20 hrs less	0.359	−0.678	−0.903**
	(0.271)	(−1.422)	(−3.230)
Interactions: VP woman X			
W 20 hrs more	−6.299*	−6.070***	−3.588***
	(−2.456)	(−6.088)	(−9.263)
W 10 hrs more	−5.828***	−2.959***	−1.859***
	(−3.369)	(−4.768)	(−5.794)
Both same hrs	−0.253	0.229	−0.0517
	(−0.144)	(0.431)	(−0.190)
W 10 hrs less	−0.615	1.693**	1.024***
	(−0.418)	(2.945)	(3.396)
W 20 hrs less	1.387	1.496*	1.825***
	(0.657)	(2.147)	(4.385)
Constant	−2.267***	−0.115	0.207*
	(−3.398)	(−0.498)	(2.216)
Number of observations	87	439	1,907
Number of individuals	29	151	642
Wald test χ^2	39.03	103.09	192.38
Wald p-value	0.000	0.000	0.000
ρ	0.166	0.235	0.077

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(11)$ for all three columns.

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.C3. RI-Regression Model of the Appropriateness Rating by Gender of the Vignette Person.

	(1) Labor market hrs	(2) Relative income	(3) Child care
Main effects: vignette factors			
VP woman (ref: VP man)	1.278	1.160	2.633
	(0.546)	(0.470)	(0.702)
Housework share	−2.452	4.274	−3.247
	(−0.357)	(0.578)	(−0.286)
Housework share^2	−6.233	−11.260	−2.549
	(−0.914)	(−1.532)	(−0.234)
Labor market hrs (ref: same)			
less hrs	−3.318		
	(−1.438)		
more hrs	−0.924		
	(−0.419)		
Relative income (ref: same)			
less income		−1.227	
		(−0.502)	
more income		2.331	
		(0.951)	
Child care share (ref: same)			
less child care			0.725
			(0.191)
more child care			−2.987
			(−0.805)
no info child care			−2.893
			(−0.780)
Interactions:			
VP woman X			
Housework share	−4.203	−4.228	−10.35
	(−0.428)	(−0.406)	(−0.664)
Housework share^2	1.309	1.995	7.770
	(0.135)	(0.193)	(0.510)
Housework share X			
less hrs	12.460		
	(1.276)		
more hrs	−1.788		
	(−0.191)		
less income		5.963	
		(0.575)	
more income		−12.920	
		(−1.239)	
less child care			−2.151
			(−0.138)
more child care			10.960

Continued on next page.

Table 2.C3 – *Continued from previous page.*

	(1) Labor market hrs	(2) Relative income	(3) Child care
			(0.705)
no info child care			13.610
			(0.884)
Housework share^2 *X*			
less hrs	−7.689		
	(−0.796)		
more hrs	3.625		
	(0.391)		
less income		−5.540	
		(−0.539)	
more income		14.060	
		(1.357)	
less child care			3.295
			(0.218)
more child care			−11.150
			(−0.728)
no info child care			−14.360
			(−0.955)
VP woman X			
less hrs	2.244		
	(0.692)		
more hrs	−0.742		
	(−0.230)		
less income		3.881	
		(1.105)	
more income		−1.673	
		(−0.479)	
less child care			−2.084
			(−0.407)
more child care			1.245
			(0.242)
no info child care			0.338
			(0.066)
VP woman X Housework share X			
less hrs	−9.904		
	(−0.726)		
more hrs	−1.515		
	(−0.111)		
less income		−18.040	
		(−1.221)	
more income		6.473	
		(0.438)	
less child care			8.876
			(0.415)

Continued on next page.

Table 2.C3 – *Continued from previous page.*

	(1) Labor market hrs	(2) Relative income	(3) Child care
more child care			−5.611
			(−0.260)
no info child care			−3.110
			(−0.146)
VP woman X Housework share^2 *X*			
less hrs	9.520		
	(0.706)		
more hrs	4.694		
	(0.349)		
less income		18.490	
		(1.266)	
more income		−6.721	
		(−0.458)	
less child care			−10.460
			(−0.496)
more child care			5.659
			(0.266)
no info child care			3.236
			(0.155)
Constant	3.140	1.136	2.534
	(1.945)	(0.655)	(0.910)
Number of observations	1,260	1,260	841
Number of individuals	422	422	401
Wald test χ^2	871	525	345
Wald p-value	0.000	0.000	0.000
ρ	0.118	0.080	0.131

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(17)$ for columns 1 and 2, and $\chi^2(23)$ for column 3.
t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.C4. RI-Regression Model of Adding Information on Exact Relative Income by Respondent Group

	(1)	(2)	(3)
	Traditional resp.	Neither nor resp.	Egalitarian resp.
Main effects: vignette factors			
VP woman (ref: VP man)	1.159	−0.371	−0.521*
	(0.772)	(−0.723)	(−2.283)
Relative income (ref: W 50% income)			
W 33% income	−1.390	−0.284	−0.382
	(−1.038)	(−0.631)	(−1.672)
W 67% income	1.501	0.102	0.161
	(1.159)	(0.213)	(0.724)
no info on rel. income	−0.176	−0.096	−0.004
	(−0.183)	(−0.241)	(−0.020)
Interactions: VP woman X			
W 33% income	1.892	0.429	0.422
	(0.892)	(0.646)	(1.325)
W 67% income	−2.222	−0.462	−0.706*
	(−1.061)	(−0.681)	(−2.236)
no info on rel. income	0.503	0.555	0.122
	(0.313)	(0.999)	(0.488)
Constant	−0.753	0.158	0.304
	(−0.867)	(0.427)	(1.876)
Number of observations	117	647	2,929
Number of individuals	39	221	984
Wald test χ^2	14.69	5.63	54.16
Wald p-value	0.040	0.583	0.000
ρ	0.133	0.144	0.062

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(7)$ for all three columns.

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.C5. RI-Regression Model on the Impact of Vignette Dimensions by Gender Vignette Person.

	(1) VP Man	(2) VP Woman	(3) Gender diff.
Main effects: vignette factors			
Married (ref: not married)	−0.095	−0.002	−0.091
	(−0.724)	(−0.015)	(−0.709)
Number/Age of children (ref: none)			
child 2 years	0.193	−0.311	0.177
	(0.915)	(−1.547)	(0.856)
child 8 years	0.235	−0.193	0.206
	(1.128)	(−0.980)	(1.016)
Housework (ref: same housework)			
40ppt less housework	1.068***	1.549***	1.067***
	(5.155)	(7.916)	(5.262)
20ppt less housework	0.405	1.008***	0.408*
	(1.920)	(5.233)	(1.979)
20ppt more housework	−0.710***	−0.823***	−0.693***
	(−3.350)	(−4.289)	(−3.351)
40ppt more housework	−1.664***	−1.667***	−1.675***
	(−8.166)	(−8.687)	(−8.409)
Labor market hrs (ref: same hours)			
20h less labor market hrs	1.427***	1.355***	1.412***
	(6.265)	(6.390)	(6.344)
10h less labor market hrs	0.487**	0.450**	0.478**
	(2.696)	(2.695)	(2.705)
10h more labor market hrs	−0.750***	−0.881***	−0.756***
	(−4.254)	(−5.041)	(−4.388)
20h more labor market hrs	−1.247***	−1.229***	−1.268***
	(−5.696)	(−5.798)	(−5.918)
Relative income (ref: 50% of hh income)			
33% of hh income	0.233	0.143	0.223
	(1.476)	(0.947)	(1.444)
67% of hh income	−0.350*	−0.582***	−0.366*
	(−2.196)	(−3.838)	(−2.351)
Share of child care (ref: same share)			
less child care	0.505*	0.116	0.546*
	(2.224)	(0.548)	(2.460)
more child care	−0.555*	−0.600**	−0.556*
	(−2.420)	(−2.759)	(−2.477)
no info child care	0.097	−0.268	0.091
	(0.428)	(−1.260)	(0.409)
VP woman (ref: VP man)			−0.138
			(−0.309)
Interactions: VP woman X			
Married			0.076
			(0.417)
child 2 years			−0.494
			(−1.689)
child 8 years			−0.407
			(−1.417)

Continued on next page.

Table 2.C5 – *Continued from previous page.*

	(1) VP Man	(2) VP Woman	(3) Gender diff.
40ppt less housework			0.491
			(1.714)
20ppt less housework			0.612*
			(2.135)
20ppt more housework			−0.122
			(−0.425)
40ppt more housework			0.016
			(0.058)
20h less labor market hrs			−0.033
			(−0.105)
10h less labor market hrs			−0.029
			(−0.117)
10h more labor market hrs			−0.121
			(−0.485)
20h more labor market hrs			0.039
			(0.126)
33% of hh income			−0.088
			(−0.402)
67% of hh income			−0.209
			(−0.948)
less child care			−0.439
			(−1.412)
more child care			−0.060
			(−0.190)
no info child care			−0.375
			(−1.203)
Constant	0.362	0.237	0.387
	(1.098)	(0.786)	(1.206)
Number of observations	636	624	1,260
Number of individuals	213	209	422
Wald test χ^2	393.95	571.57	985.75
Wald p-value	0.000	0.000	0.000
ρ	0.147	0.038	0.096

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(16)$ for columns 1 and 2, and $\chi^2(33)$ for column 3.
t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3

Do Men and Women Really Have Different Gender Role Attitudes? Experimental Insight on Gender-Specific Attitudes toward Paid and Unpaid Work in Germany.

(Sabine Düval)

Accepted manuscript of an article published in:
Social Science Research 2023, Vol. 112, 102804.
https://doi.org/10.1016/j.ssresearch.2022.102804

3 Do Men and Women Really Have Different Gender Role Attitudes? Experimental Insight on Gender-Specific Attitudes toward Paid and Unpaid Work in Germany

Abstract This article uses a novel experimental approach to measure whether men and women actually differ in their gender role attitudes. Recent research has shown that operationalizing gender role attitudes on a unidimensional scale ranging from "egalitarian" to "traditional" is problematic. Instead, their multidimensionality must to be taken into account. Similarly, an ideal measurement tool should consider that gender norms are applied conditionally, i.e., extensive information on the situational context must be provided. In this article, both preconditions are met by using a multifactorial survey experiment. The vignettes used in the survey experiment contain extensive contextual information on fictional couples' division of paid and unpaid work. In addition, the experimental variation of this information (e.g., the vignette persons' gender, the presence and age of children, and the partners' shares of paid and unpaid work) allows to disentangle the different dimensions that may influence (different) gender role attitudes of men and women. Results show no gender difference in attitudes: On average, men and women have "classical" egalitarian gender role attitudes.

3.1 Introduction

Gender role attitudes are beliefs about the appropriate roles of men and women in different areas of life (e.g., Behr et al., 2013). They result from prescriptive gender norms ("what one ought to do") that expect different behaviors based on gender (e.g., Davis and Greenstein, 2009). In the social sciences, gender role attitudes are commonly linked to various social patterns. For example, studies have found that gender role attitudes can partly explain marital satisfaction and/or instability (e.g., Minnotte et al., 2010, 2013), labor force participation of women and mothers (e.g., Kaufman and Bernhardt, 2015; Khoudja and Fleischmann, 2018), and couples' division of paid and unpaid work (for a recent overview, see, e.g., Grunow, 2019).

Most research on gender role attitudes is based on item questions for which respondents express their views on (married) women or mothers being gainfully employed (for overviews, see e.g., Davis and Greenstein, 2009; Walter, 2018a). In the context of couples' division of paid and unpaid work, historically, the belief in distinctive roles or separate spheres for men and women based on gender has been described as a traditional gender role attitude (e.g., Davis and Greenstein, 2009). Individuals with non-traditional gender role attitudes, on the other hand, do not support a division of paid and unpaid work that is based on gender in itself (e.g., Beere et al., 1984; Kroska, 2007). Much of existing literature presents the finding that "unsurprisingly" (Khoudja and Fleischmann, 2018, p. 1096) men have more traditional gender role attitudes than women.

Newer research, however, found that gender role attitudes no longer appear on a unidimensional scale ranging from "traditional" to "egalitarian". Instead, gender role attitudes should be understood as a multidimensional concept, which consists of different characteristics independent of each other (see, e.g., Grunow et al., 2018). For example, individuals may believe that in a romantic relationship both partners should contribute to the household income. At the same time, however, they may be in favor of women being the main caregivers for children. Using this newer definition of gender role attitudes, Grunow et al. (2018) still find small but consistent gender differences in gender role attitudes in Germany and in seven other European countries. The actual size of the gender difference differs depending on which concept of the (multidimensional) gender role attitudes is considered. Their findings are based on the standard item questions commonly used for research on gender role attitudes.

However, these item questions have been criticized in the past (for recent overviews, see Walter, 2018a,b). In the past, multifactorial survey experiments have been used to measure attitudes and (prescriptive) social norms (see Jasso and Opp, 1997; Opp, 2002, 2020). In multifactorial survey experiments, respondents are asked to judge short texts (vignettes) about hypothetical situations or persons with various attributes (di-

mensions); the values (levels) of these dimensions are experimentally varied across the vignettes (Auspurg and Hinz, 2015). This makes them well-suited to capture nuanced, multidimensional gender norms. In recent research, multifactorial survey experiments have been used to measure individuals' gender norms or gender role attitudes (see, e.g., Auspurg and Düval, 2022; Thébaud et al., 2021). In addition, the multifactorial survey experiment used in this article contains extensive information on the social situation of a couple; in this way, the conditionality of (social) gender norms can be better taken into account than with the standard item questions (see Auspurg and Düval, 2022).

To my knowledge, this is the first article that uses a multifactorial survey experiment to probe if there is a gender difference in prescriptive gender norms or rather in gender role attitudes toward the division of paid and unpaid work.[51] To do so, I use a multifactorial survey experiment implemented in the German population survey pairfam in wave 10 (2017/18) (Brüderl et al., 2019). In the vignette experiment, respondents should rate the appropriateness of hypothetical couples' division of paid and unpaid work based on extensive information on the couples' work-family arrangements (e.g., vignette persons' gender, employment situation, and division of housework and child care). The information presented in the vignettes randomly varied across dimensions. 2,153 respondents rated 3 vignettes each, resulting in a total of 6,370 vignette evaluations on the division of total workload of fictional couples.

This is the first study to measure gender differences in prescriptive gender norms towards the division of total workload (paid and unpaid work) by means of a multifactorial survey experiment. The measurement used here includes extensive information on both spheres of couples' paid and unpaid work, allowing for the manipulation of relevant characteristics to determine who respondents expect to do more of which task and whether respondent gender differences are observable.

3.2 Theoretical Considerations and Literature Overview

3.2.1 On the Emergence of Egalitarian Gender Role Attitudes

This article mainly wants to describe if there is a gender difference in gender role attitudes, it does not test the causal link between gender and gender role attitudes. Still, for the broader understanding of the (causal) mechanisms behind the gender difference, a short discussion of previous literature on the emergence of gender role attitudes is helpful. Changes in gender role attitudes over time are attributed to changes on the individual level (as an individual matures and gains life experience) and on the passing of historical time (through cohort replacements) (Davis and Greenstein,

[51] In this article, I mainly use the term gender role attitudes. In general, I understand gender roles attitudes as the internalization of a prescriptive gender-specific norm on the individual level (see Thébaud et al., 2021).

2009, p. 95). Seminal work on the emergence of egalitarian gender role attitudes is Bolzendahl and Myers (2004) who distinguish between exposure- and interest-based explanations.

The former is based on the notion that exposure to feminist or egalitarian ideas and situations influences a change away from traditional attitudes (Bolzendahl and Myers, 2004, p. 761). This exposure to egalitarian ideas can occur through participation in the workforce as it exposes individuals to existing gender inequalities, e.g., due to gender discrimination. At the same time, individuals might come into contact with (successfully) employed women, which should weaken stereotypical views and instead increase the understanding that women are able to perform at work, and at the same time provides examples of successful work-family balance (see Bolzendahl and Myers, 2004; Kroska and Elman, 2009). Likewise, education and socialization (e.g., through mothers' labor force participation) may expose individuals to egalitarian ideas (Bolzendahl and Myers, 2004). In contrast, interest-based explanations argue that individuals are more likely to adopt and maintain egalitarian gender role attitudes if they personally benefit from equality, for example through an elimination of wage discrimination (see Bolzendahl and Myers, 2004; Davis and Greenstein, 2009; Kroska and Elman, 2009).

Recent literature links the framework of Bolzendahl and Myers (2004) to policy feedback theory (e.g., Gangl and Ziefle, 2015; Grunow et al., 2018). In a nutshell, policy feedback theory argues that (social) policies influence (political) behavior and attitudes which in turn feeds back into the policy-making process (see Campbell, 2012). Gangl and Ziefle (2015) found for Germany that exposure to newly introduced work-family policies (i.e., the introduction of a new parental leave program) has norm setting effects insofar, that a longer parental leave also changes individuals normative expectations toward mothers' responsibility for child care. Knight and Brinton (2017) link the diversity of gender role attitudes in Europe back to the diverse policies adopted. Grunow et al. (2018, p. 47) argue that family policies change individuals' interests, beliefs, and attitudes about gender, which in turn should influence the policymaking process.

In sum, a complex interplay of life experiences, personal and social characteristics, social norms, and social policies shapes individuals' gender role attitudes (e.g., Davis and Greenstein, 2009; Grunow et al., 2018). These changes in gender role attitudes may lead to greater gender egalitarianism in the terms of support for women's and mothers' work-involvements (Bolzendahl and Myers, 2004), but may also (even simultaneously) influence individuals' support of the more traditional view that women should be responsible for child care and rearing (Gangl and Ziefle, 2015). This simultaneous support for women taking part in the labor market and for traditional family values was labeled as "egalitarian familism" by Knight and Brinton (2017).

Naturally, not only women should be affected by these mechanisms. Men can also be

expected to experience changes in attitudes (Bolzendahl and Myers, 2004). However, based on the existing literature, there are arguments why women's attitudes may have changed more in the direction of egalitarianism, while men's changes may have been slower and/or in a different direction (see, e.g., Bolzendahl and Myers, 2004; Grunow et al., 2018). In general, given their different interest structure, women may benefit more from equality than men as they directly profit from equality in earnings (Bolzendahl and Myers, 2004, p. 761). However, (heterosexual partnered) stay-at-home women or mothers may have an interest in income inequality, as they indirectly benefit from the higher earnings of their male partners. At the same time, (heterosexual partnered) men indirectly benefit from income equality through their female partners' income. This may affect men's attitudes toward egalitarianism (Bolzendahl and Myers, 2004). Similarly, men also benefit if their female partner cares for the child. This may shape their attitudes more toward "egalitarian familism."

Since both genders are exposed to egalitarian ideas through the labor market, socialization, and/or education, this should also increase egalitarian attitudes in both (Bolzendahl and Myers, 2004). For example, the experience of having a working mother in childhood is likely to reduce acceptance of traditional gender stereotypes. Still, women, may feel the penalties of gender inequality more acutely (e.g., especially in the work-place) which likely also increases their egalitarian attitudes more. Both genders are also affected by (changes in) social policies, which in turn set norms that should be adopted by men and women (Gangl and Ziefle, 2015). Since changes in social policies (e.g., a change in parental leave program) more directly affects women's lives (since they still are the primary care taker of a child), such a change could have a greater impact on their preferences and attitudes toward intensive mothering or rather "egalitarian familism" (Gangl and Ziefle, 2015).

3.2.2 The State of the Art

In all OECD countries, women are still predominantly responsible for housework and child care, even if they contribute to a couple's overall household income (OECD, 2017). Although the main pattern is similar, differences between countries can be observed. For example, women perform more than three-quarters of unpaid work in countries such as Italy and Portugal, while the gender gap is smallest in Norway, Denmark and Sweden (OECD, 2017). Countries such as Germany, the U.S., and the U.K. fall in between, with Germany closer to the traditional end of the spectrum than the latter two (OECD, 2017).

In the last decades, egalitarian gender role attitudes have been on the rise in nearly all OECD countries (see, e.g., OECD, 2017). However, the levels of egalitarian attitudes differ between countries. For example, based on data from the International Social Survey Programme, the OECD (2017) found that most young adults in Germany (i.e.,

the country and the age group in which the multifactorial survey experiment relevant for this article was conducted) show rather egalitarian gender role attitudes. Together with the U.S. and the U.K., Germany's gender role attitudes range between countries with very egalitarian attitudes such as Sweden, Norway, and Iceland, and countries with more traditional attitudes such as Turkey and Korea.[52] In a recent overview of nine European countries, also Aassve et al. (2014) found that in Scandinavian countries such as Norway, individuals not only divide their shares of housework more equally, but they also have—on average—more egalitarian attitudes than their counterparts in Western European countries such as Germany or Belgium.[53]

Most previous research has found a gendered level difference in gender role attitudes (e.g., Aassve et al., 2014; Arránz Becker, 2013; Halimi et al., 2018; Horne and Johnson, 2018; Meagher and Shu, 2019; Minnotte et al., 2010). Often, this gender gap is small but still notable and points toward women having (at least somewhat) more egalitarian attitudes than men. For example, Meagher and Shu (2019) found a gender difference in attitudes toward working mothers and especially toward working mothers of preschool children in the U.S. Aassve et al. (2014) found that men have more traditional gender role attitudes than women. Particularly relevant to this study are the findings of Arránz Becker (2013) and Horne and Johnson (2018), who found a gender gap using two item questions on gender role attitudes from the German Family Panel pairfam; the study in which the multifactorial survey experiment used for this article was conducted.

Most of this previous research interpreted gender role attitudes as an axis ranging from "traditional" to "egalitarian" attitudes (for an in-depth discussion, see Grunow et al., 2018). Newer research has deconstructed this unidimensional interpretation and refers to a multidimensional construct of gender role attitudes (e.g., Cotter et al., 2011; Grunow et al., 2018; Knight and Brinton, 2017; Yu and Lee, 2013). For example, Knight and Brinton (2017) found that meanwhile traditional attitudes have declined all over Europe, instead different specifications or profiles of "egalitarian" attitudes have emerged. They differentiate between three varieties of egalitarianism: 1) the already known "classical" egalitarian attitudes, 2) egalitarian familism (i.e., simultaneous support for women taking part in the labor market and for traditional family values), and 3) flexible egalitarianism (i.e., rejection of predefined roles for women, whether egalitarian or traditional; support for each woman's individual choice). Overall, their results support the previous finding that women score highest on the "classical" egalitarian profile, while men are more likely than women part of the traditional profile.

[52] Interestingly, the classifications differ by the exact gender role item question used. For example, solely based on the item "Both the man and woman should contribute to the household income", Germany's classification in comparison to the U.S. or the U.K would be more traditional than it would be if based on the item "A preschool child is likely to suffer if his or her mother works".

[53] Aassve et al. (2014) included Austria, Belgium, Bulgaria, France, Germany, Hungary, Norway, Romania, and Russia in their analysis.

Also notable is their finding that within the three specifications of "egalitarian" men are most likely to hold egalitarian familism attitudes (Knight and Brinton, 2017, p. 1505). Grunow et al. (2018) come to similar conclusions as they found broad support for a multidimensional understanding of gender role attitudes for various European countries.[54] Interestingly, they found the largest gender gap can still be observed within the "classical" egalitarian profile. This egalitarian profile is also the most widespread profile in most countries in their sample. For example, in Western Germany 52% of women and 41% of men were assigned to having egalitarian gender role attitudes. Western Germany is also the country with the most support for the traditional profile (9% of women and 12% of men). Within the multidimensional gender role attitudes, gender differences were relatively small. Consistent with Knight and Brinton (2017), Grunow et al. (2018) found slightly more support for egalitarian familism among men compared to women.

Most of this previous research is based on standard item questions like "A man's job is to earn money; a women's job is to look after the home and family," (ISSP, 2021) and "All in all, family life suffers when the woman has a full-time job" (ISSP, 2021). These item questions have been criticized in the past, e.g., because of mainly focusing on women's roles and most of them only portraying traditional work arrangements (see, e.g., Davis and Greenstein, 2009; Walter, 2018a,b). In addition, the wording of the item questions does not always allow to unambiguously interpret respondents' evaluations (see Braun et al., 1994). For example, in the often used item-question "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work" (ISSP, 2021) it is unclear, whether respondents agreement should be interpreted as concernment with the emotional needs of a child in case the mother is absent or if respondents rate the caring capabilities of a working mother (e.g., Braun et al., 1994, p. 34; Braun, 2008). Put differently, in this item question likely two different dimensions of gender role attitudes are mixed up; and therefore confounded. This makes it hard to disentangle both dimensions and interpret what respondents actually support. Comparably, the item question "All in all, family life suffers when the woman has a full-time job" (ISSP, 2021) mixes the work and family domains, making it difficult for supporters of egalitarian familism to give a consistent answer.

Multifactorial survey experiments have long been recommended and used to measure attitudes and (prescriptive) social norms (see Jasso and Opp, 1997; Opp, 2002, 2020). They allow to independently vary different dimensions that might influence individuals' attitudes on who should be responsible for earning, housework, and care work. This

[54] Overall, Grunow et al. (2018) found five different profiles (i.e., egalitarian, egalitarian essentialism, intensive parenting, moderate traditional, and traditional) that occurred in eight different European countries (i.e., the Czech Republic, Italy, the Netherlands, Poland, Spain, Sweden, Switzerland, and Western Germany).

makes them well-suited to capture nuanced, multidimensional gender norms.

In addition, multifactorial survey experiments allow to better capture the conditionality of (social) gender norms in comparison to the standard item questions (Auspurg and Düval, 2022). For example, in the item question "Men should participate in housework to the same extent as women" there is no information on the employment status of the "men" and "women." However, this information is probably crucial to rate the situation (assuming that a (social) gender norm is conditionally applied). Auspurg and Düval (2022) argue that respondents fill in this missing information on the social context themselves when evaluating the item questions. To do so, respondents likely base their imputations on their own experiences of the world. In other words, most item question on gender norms likely measure descriptive gender norms ("what is most likely the case") instead of the intended prescriptive gender norms ("what should be the case", see Auspurg and Düval, 2022). Vignettes contain much more information on the social context of the (hypothetical) situation, person or couple to be judged. This ensures that respondents do not need to impute information on their own to apply a (conditional) gender norm (Auspurg and Düval, 2022). By means of interaction terms, it is also possible to test conflicting normative expectations, e.g., gender deviance neutralization expects women to adhere to housework and child care in particular if they contribute more to the household income or work more in the labor market than their male partners (Auspurg and Düval, 2022).

Recent research uses multifactorial survey experiments to measure individuals' gender norms or gender role attitudes (see, e.g., Auspurg and Düval, 2022; Schulz, 2021; Thébaud et al., 2021). Some previous studies have examined gender-specific preferences or attitudes toward the division housework work by using (multifactorial) survey experiments (see, e.g., Auspurg et al., 2017b; Carriero and Todesco, 2017; Pedulla and Thébaud, 2015; Schulz, 2021). These experiments focused on attitudes toward an appropriate share of housework (see, e.g., Auspurg and Düval, 2022; Schulz, 2021), gendered beliefs about household responsibility (Thébaud et al., 2021), and perceived fairness and preferences in context to division of work (see, e.g., Auspurg et al., 2017b; Carriero and Todesco, 2017; Pedulla and Thébaud, 2015). The various experimental approaches differ not only in their research objectives and experimental set ups, but also in the cultural context in which they were conducted (i.e., the U.S., the U.K., Italy, and Germany). Interestingly, the common denominator is that in an experimental setting, when extensive information on a couple's work-family arrangement is presented, men and women do not differ in their judgments of a fair or preferred division of work. Yet norms are conceptually different from fairness or preferences (Hedström and Bearman, 2009). Despite tastes for egalitarian arrangements, individuals might follow normative views they have internalized or fulfill them anyway to avoid uneasiness and social sanctions (Ridgeway, 2011).

To my knowledge, this is the first study that explicitly looks at gender differences in prescriptive gender norms ("what should be the case") or rather gender role attitudes in context of couples' division of their total workload (i.e., housework, child care, and paid work) by using a multifactorial survey experiment. All in all, there are good arguments for why women should have more "classical" egalitarian attitudes than men, while men can be expected to (still) have more traditional attitudes than women. However, also men's attitudes likely have changed in the last decades. Especially in regard to employed women and mothers, it can be expected that men's and women's attitudes have narrowed. However, differences in attitudes toward parenting may still be expected. It is not clear from the existing literature whether egalitarian familism is more popular among men or women, or whether there is any difference at all. This article uses a novel, experimental approach to probe if men and women (still) differ in their gender role attitudes.

First, the multifactorial survey experiment allows to probe if there is an overall gender effect insofar that men and women differ in their overall attitudes toward a couple's appropriate division of work (housework, paid work, and child care). Second, as the different dimensions of gender role attitudes can be experimentally varied independently from another, the multifactorial survey experiment allows to test if men and women base their evaluations of an appropriate share of work on different dimensions. Disentangling the different dimensions allows to examine if women and men support an egalitarian division of work across all dimensions (i.e., the "classical" egalitarian attitude). At the same time it can be examined whether men still have more traditional attitudes (i.e., wanting women to do most of the housework); or if meanwhile men and women have multidimensional attitudes, like supporting an equal division of paid work, while still supporting mothers' responsibility for child care (i.e., egalitarian familism).

3.3 Data and Experimental Method

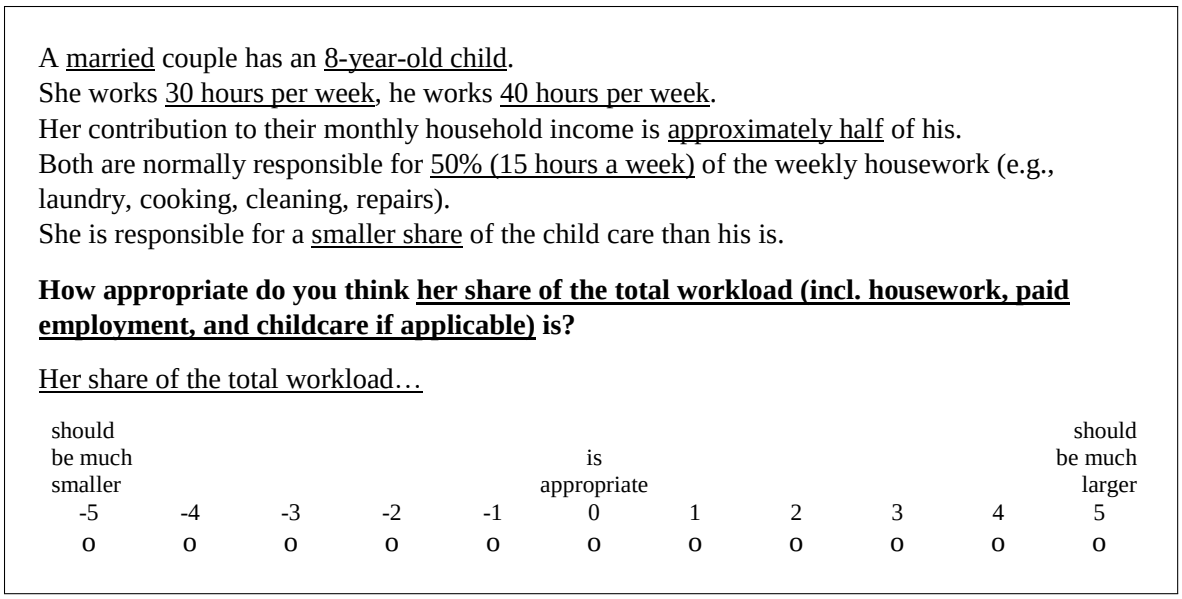
3.3.1 Data and the Respondent Sample

In 2017/18, the 10th wave of the German Family Panel (pairfam) (Brüderl et al., 2019) included a multifactorial survey experiment. Pairfam is an annual multi-actor survey first launched in 2008/09 that collects data from a nationwide, randomly selected sample of originally 12,402 respondents from three birth cohorts: 1991-1993, 1981-1983, and 1971-1973. A detailed description of the study can be found in Huinink et al. (2011).

Of the 4,750 respondents that participated in wave 10, a randomly selected subsample of 2,423 respondents should answer the multifactorial survey experiment relevant for this article.[55] I restricted the sample as follows: First, I excluded 81 respondents

[55] A total of 3,467 respondents was allocated to the experimental split on *total workload*. Of those,

who did not provide any answer on the vignette module. Second, to follow standard procedures in multifactorial survey research, I kept only evaluations of respondents with plausible response time measures in the sample:[56] In a first step, I excluded all respondents with time values above the 99th percentile and under the first percentile, followed by the exclusion of all respondents whose time values were two standard deviations above or under the mean (Mayerl and Urban, 2008; Sauer et al., 2011). Third, I removed one respondent with no information on gender from the sample. Fourth, I also excluded all respondents in a same-sex relationship at the time of the interview.[57] The final analysis sample consists of 2,153 respondents who evaluated a total of 6,370 vignettes.


A married couple has an 8-year-old child.
She works 30 hours per week, he works 40 hours per week.
Her contribution to their monthly household income is approximately half of his.
Both are normally responsible for 50% (15 hours a week) of the weekly housework (e.g., laundry, cooking, cleaning, repairs).
She is responsible for a smaller share of the child care than his is.

How appropriate do you think her share of the total workload (incl. housework, paid employment, and childcare if applicable) is?

Her share of the total workload…

should be much smaller					is appropriate					should be much larger
-5	-4	-3	-2	-1	0	1	2	3	4	5
o	o	o	o	o	o	o	o	o	o	o

Figure 3.1. Example Vignette, Manipulated Dimensions Underlined

Note: pairfam Group (2019). For the wording of the original German vignettes, see Düval and Auspurg (2020).

3.3.2 Measuring Gender Role Attitudes

The vignette module contained of three different hypothetical scenarios, called *vignettes* that described the division of work in hypothetical heterosexual couples. Respondents should evaluate whether the described share of the total workload was appropriate (*0*

2,423 respondents answered vignettes of the full-information condition relevant for this article. See Düval and Auspurg (2020) for more information on the experimental setup.

[56] This step was taken to exclude possible "satisficers" who answer quickly without giving a valid response. Keeping all 166 respondents with implausible response time measures leads to comparable results.

[57] Non-heterosexual respondents are likely to present specific gender role attitudes; however, there are not enough cases to run separate analyses on this sub-sample. Including these 19 respondents does not alter the results.

on the scale), or if one of the two partners should do more (*+5*) or less (*-5*) on an 11-point scale (see Figure 3.1 for an example of a vignette).

All scenarios describe the work and family life of fictional couples. Across the vignette scenarios, several attributes called dimensions were experimentally manipulated in their characteristics (levels). Each vignette consisted of seven dimensions: (1) couple's marital status; (2) the presence and age of children in the household; (3) the distribution of child care between partners; (4) the share of housework done by both partners; (5) the male partner's paid working hours; (6) the female partner's paid working hours; and (7) the relative earnings of both partners. Each of the seven dimensions had between two and five levels (see Table 3.1 for dimensions and levels).

The full set of scenarios, called the *vignette universe* (i.e., all possible combinations of dimension levels; see Auspurg and Hinz, 2015) spans all 7,680 possible combinations of the dimension levels.[58] I selected a sample of 750 vignettes from this vignette universe.[59] The resulting sample reflects both orthogonality and level balance, which ensures that each level occurs with approximately equal frequency (for details: Atzmüller and Steiner, 2010; Auspurg and Hinz, 2015).

In a between-respondent split, I prompted roughly half of the respondents to evaluate the housework share of the *female* vignette person [VP], while the other half evaluated the housework share of the *male* VP by random assignment. Both splits contained the same fraction of 750 vignettes, so that information on male and female VPs was symmetrical. Respondents were randomly allocated to one of the splits and, within the split, into subsamples of three vignettes (from the fraction of 750 vignettes). The order of the three scenarios was randomized for each respondent to neutralize possible vignette order effects (Auspurg and Jäckle, 2017).

Respondents answered the posed vignettes in a self-completed module (CASI) of the questionnaire. As vignettes are likely to be better understood by respondents when reading them themselves, and because this mode is known to reduce social desirability bias, this is the preferred mode for multifactorial survey experiments (Auspurg and Hinz, 2015). Prior to the implementation in pairfam, Düval and Auspurg (2020) conducted a number of pretests that suggested respondents coped well with the vignette questions and the level of complexity.

3.3.3 Analysis Strategy

As is standard in multifactorial survey research, this article uses linear regression models to predict vignette evaluations (Auspurg and Hinz, 2015). More specifically, I

[58] Note that the combination of all dimension levels in this article is only 3,240. To allow for comparisons between different experimental conditions, the full vignette universe was used for all splits.

[59] To ensure that there are no confoundings between the main effects of the vignette dimensions, the two-way interactions, and the three-way interactions, I used a D-efficient sampling technique.

Table 3.1. Vignette Implementation: Dimension and Level Overview

	Dimensions	Levels
1	Marital status	Unmarried/married
2	Presence and age of children	No children/2 years/8 years
3	Share of child care (man/woman)[a]	Larger share than partner/smaller share than partner/same share as partner/no info
4	Share of housework per week (man/woman)[a]	70% (21h)/60% (18h)/50% (15h)/40% (12h)/30% (9h)
5	Labor market hours per week (man)	40/30/20 hours
6	Labor market hours per week (woman)	40/30/20 hours
7	Relative earnings (man/woman)[a]	Twice as much as partner/half as much as partner/same as partner

Note: [a] Approximately half of the respondents were informed about the relative share of work/earnings of the male/female VP.

estimated regressions with random intercepts [RI] to account for the hierarchical data structure, as each respondent evaluated three separate vignettes.[60] I conducted all analyses using the statistical software Stata (StataCorp, 2019).

In order to represent the evaluation of the female VP's share of total workload and to analyze evaluations for both VPs' genders simultaneously, I generated a combined response scale by multiplying evaluations of male character vignettes by -1, while the evaluations of female character vignettes stayed the same. The resulting evaluation allows for simpler interpretations: Positive (negative) values represent the attitude that the female vignette person should do more (less) of the total workload than indicated in the vignette, while a zero-effect indicates that her total workload is considered appropriate.[61] I included all vignette dimensions in the models, re-coded to indicate the female vignette person's characteristics relative to the male vignette person's. See Table 3.A1 in Appendix A for the resulting vignette dimensions and levels.

[60] For more information on the use of random intercept [RI] models in factorial survey research, see Auspurg and Hinz (2015).

[61] To evaluate whether the analysis of the VP's share of total workload leads to interpretations similar to those reported, a binary variable that indicates whether the VP is female was included into an RI regression model on the VP's share of the total workload. Comparable to the results reported, a small, negative effect of the VP's gender on the appropriateness of the VP's total workload (for female respondents: β = -0.61, $p < 0.001$; for male respondents: β = -0.59, $p < 0.001$) was found.

3.4 Analysis

3.4.1 Descriptive Results

The final analysis sample consists of 2,153 respondents. Slightly less than half (46.8%) of the sample are men, the average age is 36.1 years (standard deviation [$s.d.$] = 8.0), and 77.1% of respondents are married or in a partnership. For an overview of relevant descriptive statistics, see Table 3.A1 in Appendix A.

First descriptive results indicate that respondents' evaluations of the multifactorial survey experiment are distributed across the whole response scale (see Figure 3.2).[62] The mean evaluation across gender is -0.29. As information on male and female vignette persons is symmetrical (i.e., on average, male and female VPs do the same hours of housework, paid work, and child care), an average vignette evaluation of 0 shows that respondents do not differ much in their evaluations for male and female VPs. But is there a difference by respondent gender? Results from a random intercept [RI] regression model show that on average, men and women want female VPs to do a bit less of the total workload (men: -0.27; women: -0.31). The mean difference by respondent gender is only -0.04 (p = .457).

Evaluations from men and women are approximately the same. This is a first indication that men and women do not differ in their gender role attitudes. Note, however, that these are summarized statistics for all vignettes. This can merely be interpreted as indication that there is no gender difference in men's and women's overall ratings. In the next section, a closer look into the factorial survey experiment, allows to disentangle the relevant dimensions (such as the share of housework, paid work, and care work) of a multidimensional gender role attitude. In this way, a possible gender difference can be examined in more detail.

3.4.2 Main Results

Figure 3.3 shows results from RI regression models. The left section shows a β-coefficient plot of the vignette dimensions with 95% confidence intervals. These estimates are based on two separate RI regressions by respondent gender. The right section shows a plot of the gender difference of each coefficient based on a joint RI model for men and women that includes the gender of the respondent. Interactions of the respondent gender variable with the respective vignette dimensions allow for effect comparisons. The coefficient plot was created with the Stata command *coefplot* (Jann, 2014). For detailed regression results, see Table 3.A2 (models 1-3) in Appendix A.

The effects of the individual vignette dimensions suggest that the appropriateness of the female VP's total workload depends on her relative shares of housework, child

[62] Almost all respondents used the response scale to grade their answers. Only 274 respondents marked the same answer in all three vignettes. Excluding these individuals does not affect the results.

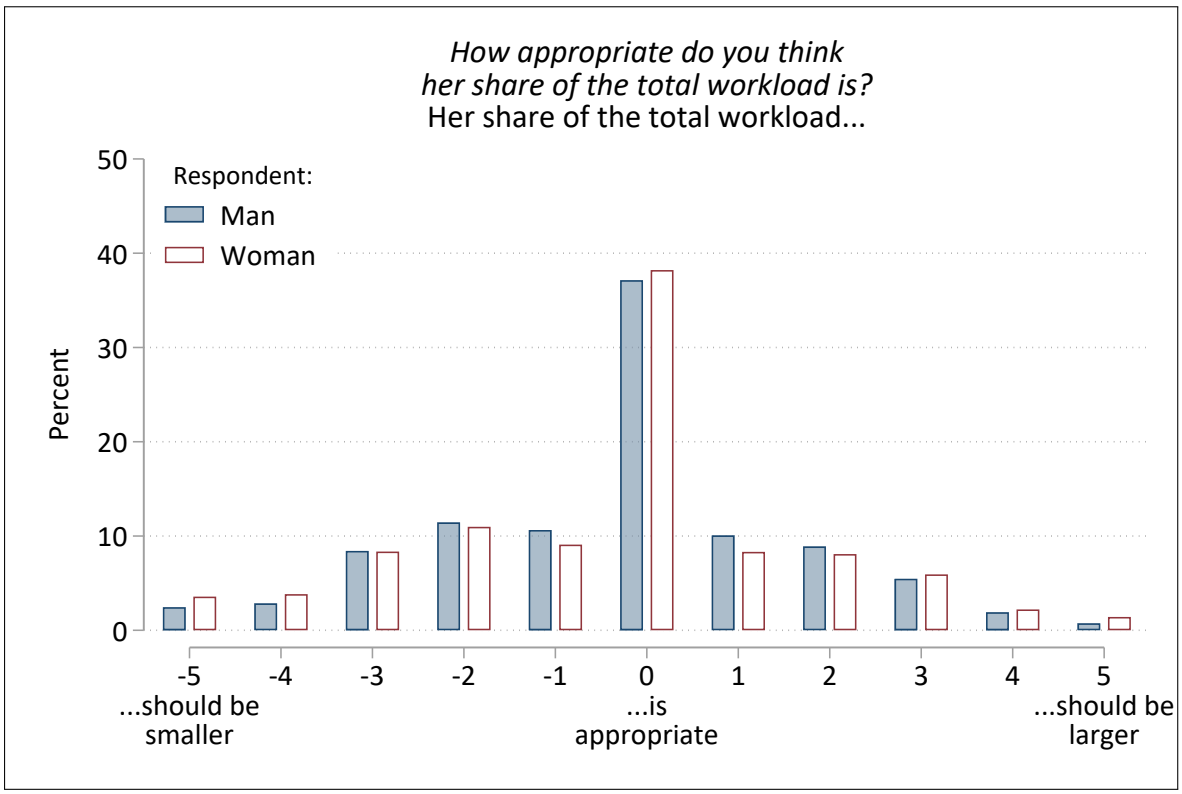


Figure 3.2. Distribution of Responses by Respondent Gender

Note: These figures show the distribution of responses to factorial survey experiment separately by respondent gender.
Number of respondents in total: 2,153, thereof 1,008 men and 1,145 women. Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

care, and paid work, and to a lesser extent also her relative earnings; marital status and presence and/or age of children do not statistically significantly influence the ratings (see left section of Figure 3.3). Overall, the coefficients hardly differ in effect size between genders, and the gender differences are also not statistically significant (see right section of Figure 3.3).

The results indicate that men and women support the classical egalitarian division of work (i.e., all contributions to the partnership in the form of working hours or income increase the appropriate share that is done by a partner, regardless of that partner's gender).[63] If men supported traditional gender norms, an additional effect of the VP gender would be expected. But the constant suggests that men rate it appropriate when both partners contribute the exact same shares of unpaid work, paid work, and earnings (i.e., all vignette dimensions in their reference category).[64]

[63] In research on distributive justice, this is also known as the rule of equity (Deutsch, 1975).

[64] A closer look into the vignette dimensions reveals the tendency that men already rate it appropriate if women do a bit less of the paid employment and earn a bit less than men (see slight asymmetry in men's coefficient for "Share of paid work" and "Relative earnings" in the left section of Figure 3.3). This might be a small hint toward traditional attitudes. However, the difference is rather small and should therefore not be overestimated.

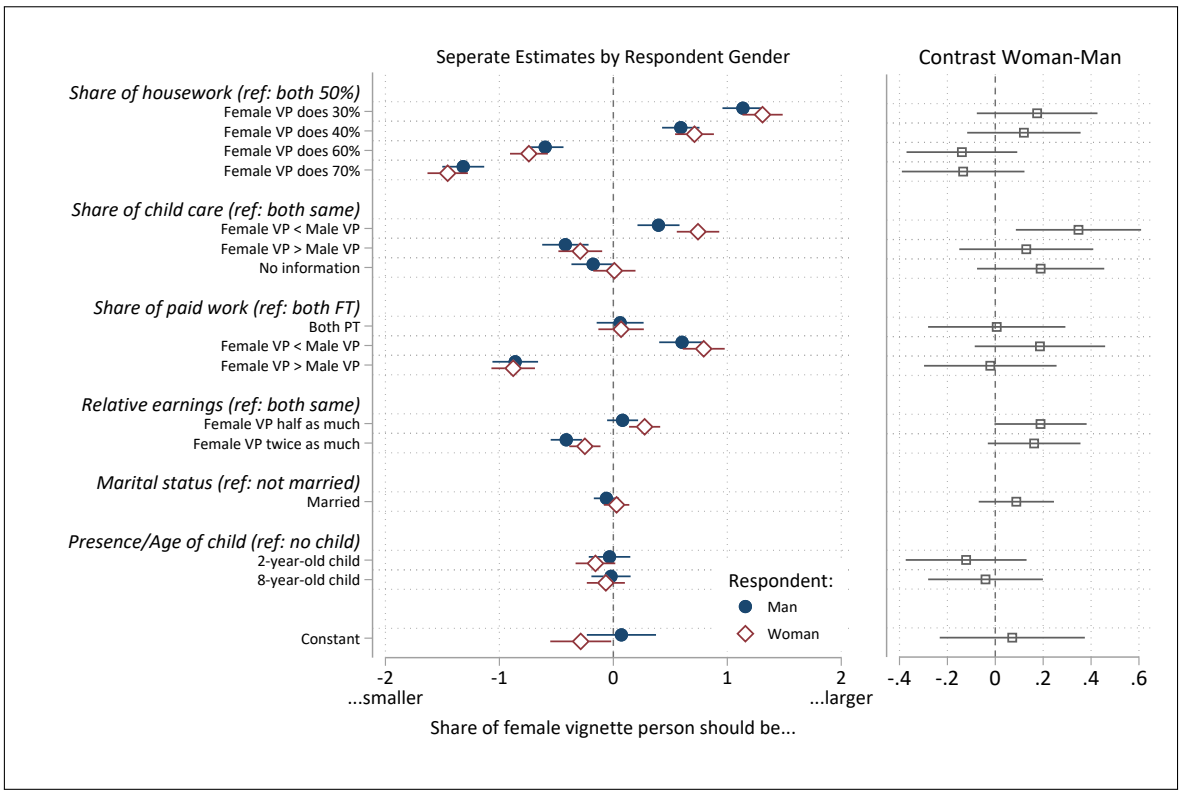


Figure 3.3. Vignette Evaluation Coefficient Plot by Respondent Gender

Note: This figure shows the results of RI regressions of the evaluation of female vignette person's (VP's) total workload on vignette dimensions. The left panel regressions were estimated separately men and women. The right panel shows the differences in effect sizes across both genders. For the regression tables, see Table 3.A2 (models 1-3) in Appendix A.
Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

Interestingly, for women is there a small asymmetry in effect sizes in cases where the female VP provides less vs. more child care than her partner. This asymmetry hints toward women's support for "intensive mothering" (i.e., egalitarian familism).[65] Here, women and men differ statistically significant from another (see the positive effect for "Female VP $<$ Male VP" on the right section of Figure 3.3). To look deeper into women's support for "intensive mothering", I included interaction terms of the female VP's share of child care with the age of the child (2-year old or 8-year old) in an additional analysis. The age of the child does not statistically significant alter the effect of child care. This means that women's support for "intensive mothering" is not dependent on the age of a child.[66]

Thus far, only linear effects of the vignette dimensions on the appropriateness of the female VP's share of the total workload have been considered. This approach

[65] In more detail, the regressions coefficients for women indicate that the amount of workload a female VP should increase her workload (β = 0.74) if she does less of the child care is greater than the amount of workload she should decrease her workload (β = -0.30) if she does more of the child care.

[66] The results are not shown here but can be requested from the author.

allows for a test of whether men and women differ on the factors that influence their ratings of an appropriate division of paid and unpaid work. However, an (additional) gender difference might emerge if interactions of the individual vignette dimensions are considered. For example, supporters of egalitarian familism might rate it appropriate if employed women (who might even do more hours in the labor market than their male partners) do also most of the child care. Individuals who hold traditional attitudes might consider it appropriate for employed women who work full-time or even more than their male partners to additionally do most of the housework. Consequently, the effects of certain vignette dimensions (e.g., hours spent on housework and shares of child care) might differ according to the level of other vignette dimensions (e.g., paid working hours and relative income). Technically speaking, I included interaction terms of the unpaid with the paid work dimensions in the RI regression models. For the regression table, see Table 3.A2 (models 4a-c) in Appendix A.

Figure 3.4 shows the appropriateness of the female VP's total workload by her share of housework (i.e., ranging from 30% to 70% of total housework). Four different employment situation are shown: (a/e) both partners work full time; (b/f) both partners work part time; (c/g) the female VP has less paid working hours than her male partner (i.e., a male breadwinner); and (d/h) the female VP has more paid working hours than her male partner (i.e., a female breadwinner). In the first row, the predicted scores of the vignette evaluation for each of the four situations are shown. In the second row, the gender difference for the corresponding situation is shown. Predicted values and contrast were calculated at the following levels of vignette dimensions: "no children", "not married", "same income", and "no information on child care".

There is no indication for a gender difference. In situations with a male breadwinner, men and women want the female VP to do more of the housework than her male partner (see Figure 3.4 c). Her share of the total workload is found to be appropriate (i.e., a predicted score of 0) if she does 60% of the housework (i.e., 18 of 30 hours per week). If she does more (less) than 60%, respondents want her share of the total workload to be smaller (larger). Also in situations in which a woman deviates from her "normative role" as homemaker (i.e., relatively more hours in paid work), both genders want her to be responsible for relatively less hours of housework (see Figure 3.4 d). In these situations, both men and women find it appropriate for the female VP to be responsible for 30% of the housework. If both partners work full time, a housework share of 50% each is considered appropriate by men and women (Figure 3.4 a). These results are evidence for wide support for the classical egalitarian gender role attitudes. There is no statistically significant gender difference (see Figure 3.4 e-h).[67]

[67] The only statistically significant difference can be seen in Figure 3.4 f: Women want the female VP's total workload to be relatively smaller when both partners work part time and the female VP does 60% of the housework. The difference is, however, rather small and in the range of what can be defined as statistical significance expected by chance (i.e., 1 of 20 situations).

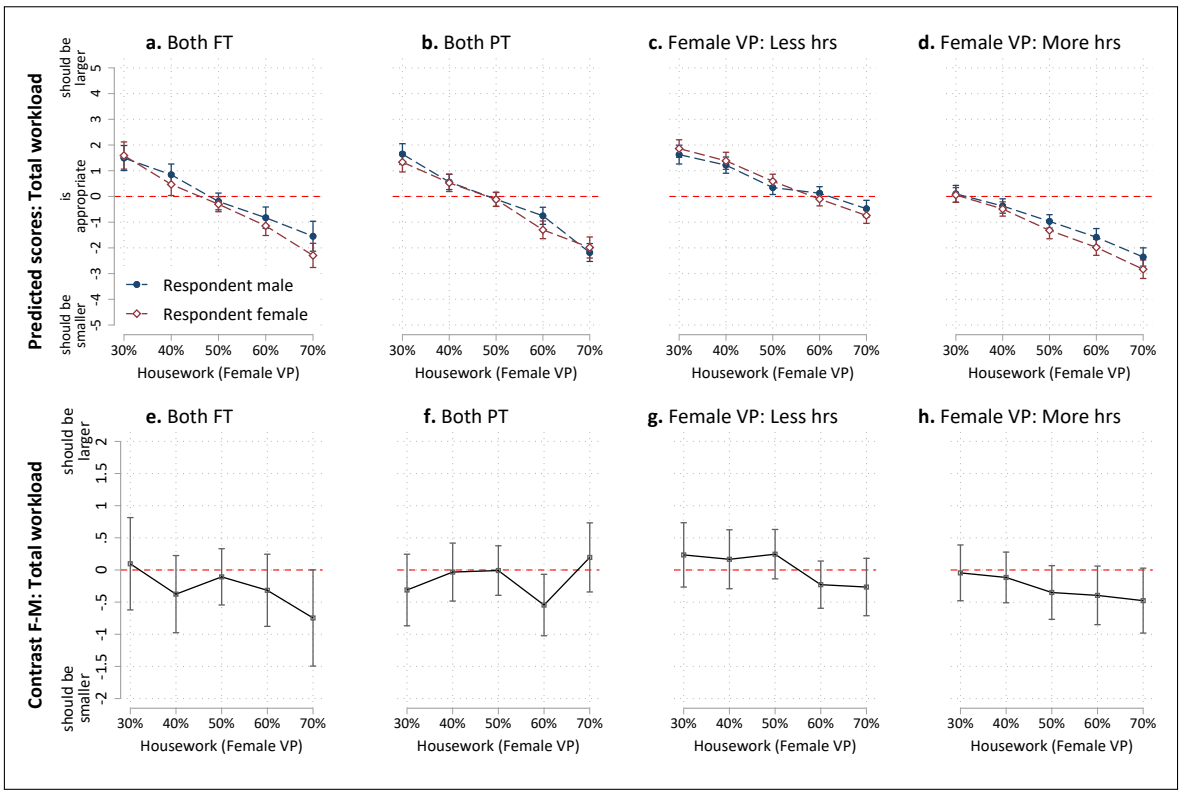


Figure 3.4. Appropriateness of the Female VP's Total Workload by Housework Share and Paid Employment

Note: Figure 3.4 shows the appropriateness of the female VP's total workload by housework share (i.e., ranging from 30% to 70% of total housework) and paid employment. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's paid working hours: (a/e) both partners work full time (FT); (b/f) both partners work part time (PT); (c/g) the female VP has less paid working hours than her male partner; and (d/h) the female VP has more paid working hours than her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.A2 (models 4a-c) in Appendix A.
Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

Figure 3.5 shows the effects of the female VP's share of child care (i.e., less than, the same as, or more than the male partner's share) in the four relevant employment situations. Predicted values and contrast are calculated at the following levels of vignette dimensions: "child: 2 years", "not married", "same income", and "housework 50%".

The overall tendency is the same as for housework: Being responsible for a larger share of child care allows for a smaller share of the total workload to be found appropriate in all four employment situations (see Figure 3.5 a-d). Again, VPs with a plus (minus) in paid working hours should do a smaller (larger) share of child care for an appropriate total workload (i.e., a predicted score of 0). There are no indications for interaction effects. Especially, there is no evidence that women are expected to do more of the child care, independent of their employment situation. Differences by

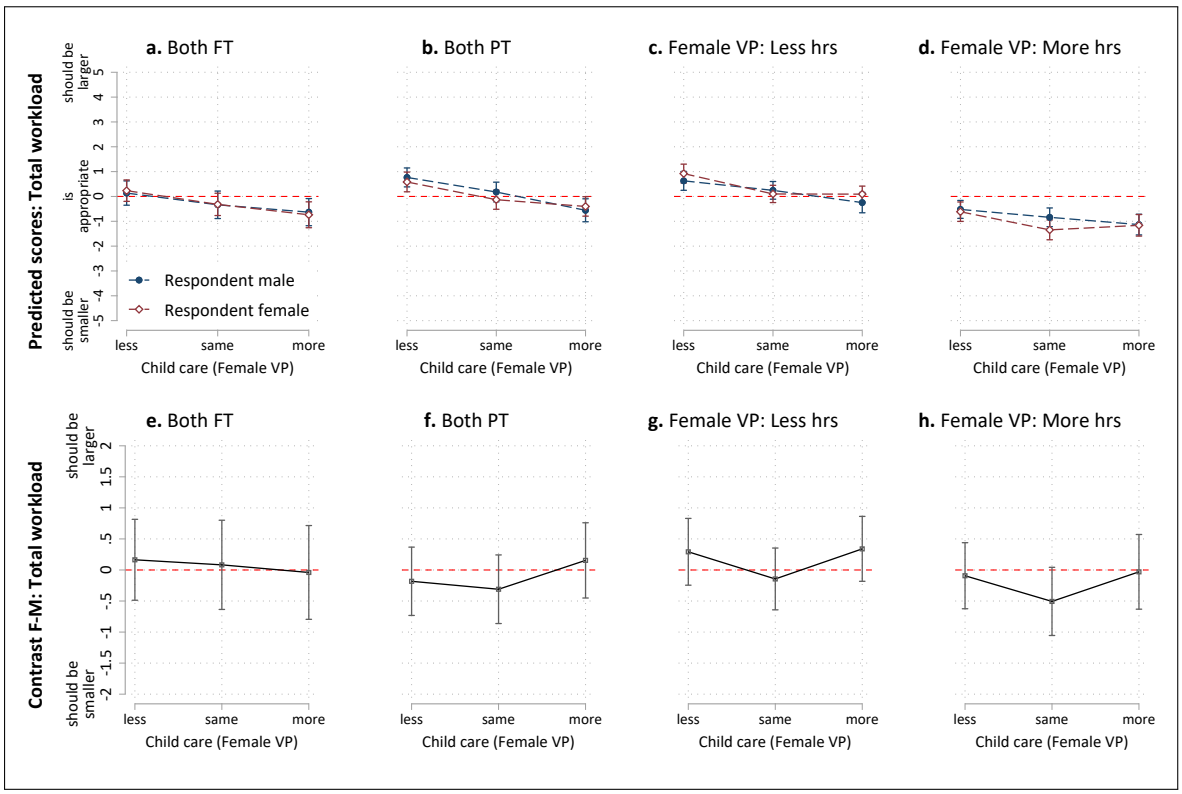


Figure 3.5. Appropriateness of the Female VP's Total Workload by Share of Child Care and Paid Employment

Note: Figure 3.5 shows the appropriateness of the female VP's total workload by share of child care (i.e., less, same, or more than the male partner) and paid employment. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's paid working hours: (a/e) both partners work full time (FT); (b/f) both partners work part time (PT); (c/g) the female VP has less paid working hours than her male partner; and (d/h) the female VP has more paid working hours than her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.A2 (models 4a-c) in Appendix A.
Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

respondent gender are not statistically significant (see Figure 3.5 e-h).

Research found, that men's and women's gender role attitudes particularly differ toward working mothers of preschool children (see, e.g., Meagher and Shu 2019). To pay special attention to preschool children, Figure 3.5 shows the appropriateness of the total workload in case a *2-year-old child* is present. In Appendix A, Figure 3.A3 shows similar results for an *8-year-old child*. The results are comparable to the ones presented here. Overall, there is no indication that the age of the child matters for a gender differences in gender role attitudes.

Thus far, only the effects of unpaid work in the four different employment situations have been considered. However, the effects of housework or child care on the appropriateness of the woman's total workload might differ by the couple's specific income

(i.e., both earn the same amount, woman earns half as much, and woman earns twice as much). Figures 3.A1 and 3.A2 in Appendix A show that being responsible for a larger share of housework or child care allows a smaller share of the total workload to be found appropriate in all three earnings situations. No interaction effects are detectable. Overall, differences between income situations are rather small, resulting in predicted scores of the total workload being more dependent on housework hours and share of child care than on income situation. Again, results for men and women are comparable; there are no statistically significant differences by respondent gender.

3.4.3 Validity and Robustness Checks

The factorial survey experiment implemented here allows for a differentiation between traditional and non-traditional gender role attitudes. With the experimental set-up it was possible to test whether men and women differ in their gender role attitudes. A random variation of different dimensions allowed for a more detailed examination of the multidimensional construct of gender role attitudes. This is a great advantage over classical gender role item questions. Nevertheless, this analysis was not without its limitations.

First, it could be argued that when respondents affirm that the female VP's shares of total workload should be larger, it is not known whether her total workload should increase by increasing her a) hours in paid work or b) hours in unpaid word, including housework hours and/or child care. Here, a difference between men and women might arise: For example, men might want female VPs to increase their unpaid work, while women want them to increase their time in paid work. The same patterns could be argued if respondents want a female VP do decrease her total workload.

A small additional experiment asked respondents to rate the appropriateness of the VPs' share of housework (and not their total workload). This additional experiment was included in the same pairfam wave as the original experiment in a between-respondent split. Results for this additional experiment can be seen in Figures 3.B1-3.B4. For the regression tables, see Table 3.B1 (models 4a-c) in Appendix B.[68] Results are consistent with the original findings on total workload: For example, the first row of Figure 3.B1 (in Appendix B) shows that for female VPs, a relatively high share of paid work compared to a relatively low share of paid work should lead to a reduction in the amount of housework for the share of housework to be considered appropriate. If men indeed had traditional attitudes, they would be expected to consider a division of housework appropriate in which women who do more paid work hours than their partners still do a larger share of the housework. And even if men agreed (at least

[68] As fewer respondents took part in the additional experiment than in the original experiment (384 vs. 2,153 respondents), this lead to fewer vignette evaluations (1,149 vs. 6,370 vignette evaluations). Confidence intervals are therefore larger for the additional experiment.

to some extent) that it is also appropriate for women to reduce their time spent on housework if they do more paid work, a gender gap is still expected if men indeed have more traditional attitudes than women. However, I found no gender difference, as men and women do not differ statistically significant in their views on the appropriate division of housework (see second row in Figure 3.B1 in Appendix B). In sum, both men and women have classical egalitarian attitudes toward an appropriate division of housework.

This bolsters the idea that men and women also adhere egalitarian ideas when evaluating the appropriateness of the total workload: If women (men) do too little unpaid work or too little paid work, women (men) should increase their total workload, either by spending more time in unpaid work or more time in paid work. This can also be seen in Figure 3.3 (see section 3.4.2), where an appropriate share of the total workload can be achieved either by increasing/decreasing the share of unpaid work or by increasing/decreasing the share of paid work. Again, no gender differences are found.

Second, the external validity of the estimated results could be limited due to the relatively young age of the pairfam sample: At the time of the interview in 2017/18, respondents were between 23 and 47 years old. It could be argued that there is no difference between the genders only because of the young age of the sample. As the pairfam survey also includes a classical item battery on gender role attitudes, this allows to compare the results over measurement instrument. I made sure that the analyses on the item questions and the analyses on the factorial survey experiment contain the same respondents. This ensures that any possible difference in the results is not based on different sample compositions. The four item questions of the pairfam gender role item battery are posed by the interviewer (CAPI mode) directly at the beginning of the questionnaire, while the factorial survey experiment was one of the last modules included. The additive index of all four items shows a relatively low reliability, with a Cronbach's Alpha of only .51. For this reason, only the results of the item that represented attitudes toward the division of paid and unpaid work most closely ("Men should participate in housework to the same extent as women") are presented here, in line with previous research (e.g., also with pairfam data, Nitsche and Grunow, 2016).[69] The results for the item question show that most respondents do not have traditional gender role attitudes. The average agreement is slightly higher for women ($mean$ = 4.3 ($s.d.$ = 0.9)) than for men ($mean$ = 4.1 ($s.d.$ = 0.9)). According to a t-test, this difference is statistically significant ($t(2151)$ = -3.9, $p < .001$). This is in line with previous research on gender differences in gender role attitudes (see, e.g., Grunow et al., 2018).

Third, research on attitudes in surveys is known to be prone to a social desirability bias. Respondents may have felt pressured to answer the factorial survey experiment in

[69] The results of the additive index are comparable to the single item question.

a socially desirable way. Methodological research suggests that factorial survey experiments are less prone to social desirability than item questions (see, e.g., Auspurg and Hinz, 2015). In the case of pairfam, the factorial survey experiment was answered in the self-administered module (CASI), in which interviewers cannot see the responses. The item questions were posed by an interviewer. If social desirability is present in the data, it should affect the item questions to a larger extent. Therefore, it is unlikely that social desirability in answers to the factorial survey experiment explains the non-existing gender difference or the overall support for an egalitarian distribution of work. Nevertheless, one could still argue that female interviewers provoked more egalitarian answers than male interviewers, as female interviewers are generally expected to hold more egalitarian attitudes than male interviewers (see, e.g., Klein and Kühhirt, 2010). In an additional check, no statistically significant difference in men's or women's answers to the factorial survey experiment by gender of the interviewer was found ($p > 0.05$).

Finally, I conducted a number of additional validity tests. The results presented here were checked for dependency on a specific model by re-running all analyses with ordinary least square regressions and cluster-robust standard errors. For the regression tables, see Table 3.B2 in Appendix B. The results of this secondary analysis were comparable to those reported in section 3.4.2. As already mentioned in section 3.3.1, further analyses ensured that the results were not artificially influenced by "satisficers" (i.e., respondents who answer too quickly without giving valid responses) or by sample restrictions.[70] Furthermore, I considered that factorial survey experiments are a rather complex format not familiar to pairfam respondents; therefore, respondents might have needed some time to adjust to the new survey format. For this reason, I excluded each respondent's first vignette, (i.e., often considered a "training vignette", see Auspurg and Hinz, 2015). The results of analyses run without this first vignette are also comparable to the initial results. For the full regression tables, see Table 3.B3 in Appendix B.

3.5 Summary

I summarize the main findings of this analysis in three points. First, I found no evidence of a general difference between genders in attitudes toward the appropriate division of workload (i.e., housework, child care, and paid work).

Second, I found evidence that men and women predominantly have attitudes that can be assigned to the "classical" egalitarian profile. When evaluating an appropriate division of paid and unpaid work, both genders differentiate between different inputs (i.e., shares of housework, child care, paid work, and relative earnings) and count them against each other to rate the appropriate division of workload. Contrary to my

[70] As it is known from research that respondents who answer too quickly are more prone to method effects, like, e.g., dimension order effects (see, e.g., Düval and Hinz, 2020).

expectations, men do not want women to do more unpaid work simply because of their gender. This argues against a traditional attitude among men. I found some evidence that women support "intensive mothering" (i.e., egalitarian familism) to some extent. Interestingly, men do not.

Third, I found no evidence for possible interaction effects of the vignette dimensions. Instead, linear effects of the amount of unpaid work on the appropriateness of the total workload in a wide range of different employment and earnings situations is evident. Taking on more housework or child care allows for a reduction in overall workload, for both men and women. At the same time, women who work relatively longer hours in paid employment or earn relatively more than their partners require a smaller share of unpaid work in order for their total workload to be evaluated as appropriate. Again, this is clear evidence that men and women, on average, have classical egalitarian attitudes toward gender roles.

3.6 Discussion

This contribution discussed whether men and women differ in their gender role attitudes. Based on the results of this article, men and women predominantly hold "classical" egalitarian gender role attitudes. This is in line with previous research by Grunow et al. (2018), Knight and Brinton (2017) and many more. I found some evidence that women support "intensive mothering" to some extent. This is in contrast with previous findings that primarily assigned men to the "egalitarian familism" profile (for previous findings, see Grunow et al., 2018; Knight and Brinton, 2017). In further analyses, that included interaction terms of the different dimensions of the multidimensional gender role attitudes, I did not find any support for "intensive mothering" in either men or women.

Overall, the results clearly showed that there is no difference in gender role attitudes between men and women. This is in contrast to most of the previous literature that used item questions to measure gender role attitudes (Aassve et al., 2014; Arránz Becker, 2013; Grunow et al., 2018; Halimi et al., 2018; Knight and Brinton, 2017, and many more). As suggested by previous research (Jasso and Opp, 1997; Opp, 2002, 2020), I used a multifactorial survey experiment to measure social gender norms. Respondents of the large scale German family panel pairfam received information on both partners' hours of housework and paid work, shares of child care, and relative earnings. Additionally, the vignettes contained information on couple characteristics (e.g., marital status and the presence and age of children in the household). This experimental approach allowed me to disentangle the different dimensions of the multidimensional construct that are gender role attitudes. In addition, I provided respondents with extensive information on the work-family situation of couples described in the vignettes.

This allows to better capture the conditionality of gender norms and is particularly important when comparing gender role attitudes across different respondent groups (Auspurg and Düval, 2022).

Why did I find no gender difference but previous research did? The aim of this study was mainly to describe if there is a difference in men's and women's prescriptive gender norms or rather gender role attitudes. As argued, a multifactorial survey experiment is well-suited to measure social gender norms. It is up to future research to test why the results of the multifactorial survey experiment in this article differ from previous research on gender differences in gender role attitudes. Auspurg and Düval (2022) argue that differences in gender norms between respondent groups might be caused by different experiences these groups have about of the world. This might lead to different interpretations of the standard item question; and may result in differences in measured attitudes. Empirically, this could be tested by conducting a factorial survey experiment on descriptive gender norms of men and women.

Overall, the results of this article are in line with previous experimental research on gender differences in related concepts such as preferences or fairness evaluations (e.g., Auspurg et al., 2017b; Carriero and Todesco, 2017; Schulz, 2021). To date, only few experimental studies on gender-specific preferences, fairness evaluations or even gender norms have been conducted in selected (mostly European) countries that are considered to be neither particularly traditional nor particularly egalitarian. What all these experiments have in common is that men and women do not differ when stating their attitudes, preferences or fairness evaluations in regard to a couples' division of paid and unpaid work. It would be interesting to see if this is also the case at the very traditional end of the spectrum, or if there is still an active social norm on gender-segregated divisions of work in countries not yet covered by survey experiments.

As any investigation, this study also has its limitations. First, responding to the factorial survey experiment relevant for this article is a rather demanding task. Respondents must understand various information and possibly calculate and compare amounts of paid work, housework, and earnings against to decide whether a certain division of paid and unpaid work is appropriate. This could lead to cognitive overload or fatigue effects. Each respondent answered 3 vignettes with 7 dimensions, which is in the range of what is recommended by methodological research on factorial survey experiments to avoid such biases (see, e.g., Auspurg and Hinz, 2015; Auspurg and Jäckle, 2017). As the respondents of the pairfam sample are rather young (between 23 and 47), this should further avoid possible biases. Nevertheless, respondents might have needed some time to adjust to the new survey mode. Excluding a "training vignette" did not alter the results. Furthermore, before the implementation in the pairfam survey, a pretest ensured that respondents coped well with the information presented in the experiment.

Second, the external validity of the estimated results could be limited due to the relatively young age of the pairfam sample. It is unclear whether older respondents might (still) show a gender difference when measured with a factorial survey experiment. It is noteworthy, however, that even for the relatively young (and therefore possibly relatively progressive) pairfam sample, an existing gender difference measured with the item questions was found, that then disappears when measured with the factorial survey experiment.

Third, only one aspect of gender role attitudes is measured in the factorial survey experiment: the division of paid and unpaid work. Other relevant aspects (e.g., public roles or gendered power balances in paid work) are not measured. This study therefore only covers a small part of gender role attitudes. Some authors argue that men and women differ especially in their attitudes toward working mothers and the wellbeing of their (preschool) children (e.g., Meagher and Shu, 2019). Although the present analyses did not consider the wellbeing of children, the results suggest that the presence or age of children does not influence the perceived appropriateness of a couple's division of paid and unpaid work. Still, some results of this article hinted toward women's support for intensive mothering. For future research, it is probably fruitful to take a deeper look into gender differences in regard to child care and possible negative consequences for children if their parents and especially their mothers work.

Fourth, this multifactorial survey experiment only provided information on the amount of housework and child care completed by men and women. Research has argued that different housework and child care tasks are specifically ascribed to men or women (e.g., Schneider, 2012). Future research could expand these results to investigate whether they also hold when differentiating between different specific tasks.

Finally, an intersection of racial and gender attitudes is argued to be the way to go when analyzing attitudes on inequalities or discrimination (e.g., Scarborough et al., 2021). The multifactorial survey experiment this article is based on only differentiates between hypothetical cis men and cis women in heterosexual relationships. There was no information on the race or the migration background of the couples. Moreover, the findings of this article on respondent gender differences are applicable only to cisgender individuals who are not in homosexual relationships.[71] As they are highly relevant in this area of research (e.g., Scarborough and Risman, 2017), to explore the gender role attitudes of individuals that are part of the LGBTQ communities, separate analyses would be necessary. However, in the pairfam survey, only one respondent identifies as transgender. Only 19 are in homosexual relationships. To avoid marginalizing the voices of LGBTQ communities, I decided to not include these individuals in the main analyses, as they would get lost in the mass of responses from heterosexual, cisgender individuals.

[71] The pairfam survey only differentiates between heterosexual and homosexual relationships.

In sum, the results show that meanwhile, at least on average, men and also women have (classical) egalitarian attitudes toward couple's division of paid and unpaid work. They do not differ in their respective gender role attitudes here. Gender inequalities are a complex interrelation of individual, interactional, and macro dimensions (e.g., Scarborough and Risman, 2017). Therefore, it is short-sighted to say that reducing traditional gender norms is the (only) key factor to reduce gender inequalities. Policies (e.g., federal regulations on parental leave), personal preferences, and disparities in economic resources of men and women arguable also play a significant role in forming differences in men's and women's shares of paid and unpaid work (e.g., Gangl and Ziefle, 2015; Scarborough and Risman, 2017). However, cultural processes (e.g., a gender norm on who should be responsible for paid and unpaid work) also are part of the dynamics of gender inequalities (Scarborough and Risman, 2017). So it is noteworthy that at least according to men's and women's attitudes, equality of outcome (i.e., both partners complete approximately the same amounts of paid and unpaid work) could be reached. Still, further research is needed to shine light on other cultural processes, like (persistent) norms on the ideal (male) worker and intensive mothering.

Appendices

Appendix A: Main Results

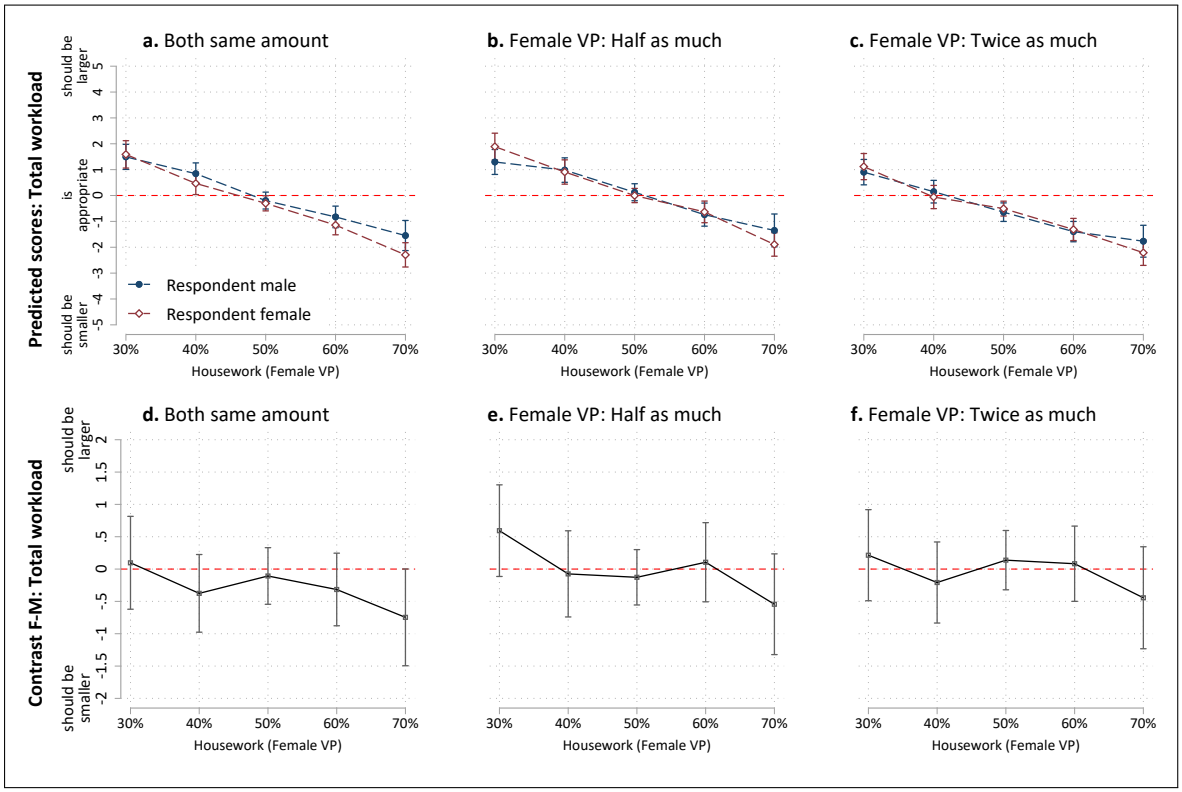


Figure 3.A1. Appropriateness of the Female VP's Total Workload by Housework Share and Relative Earnings

Note: Figure 3.A1 shows the appropriateness of the female VP's total workload by housework share (i.e., ranging from 30% to 70% of total housework) and relative earnings. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's relative earnings: (a/d) both partners earn the same amount; (b/e) female VP earns half as much as her male partner; (c/f) female VP earns twice as much as her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.A2 (models 4a-c) in Appendix A.
Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

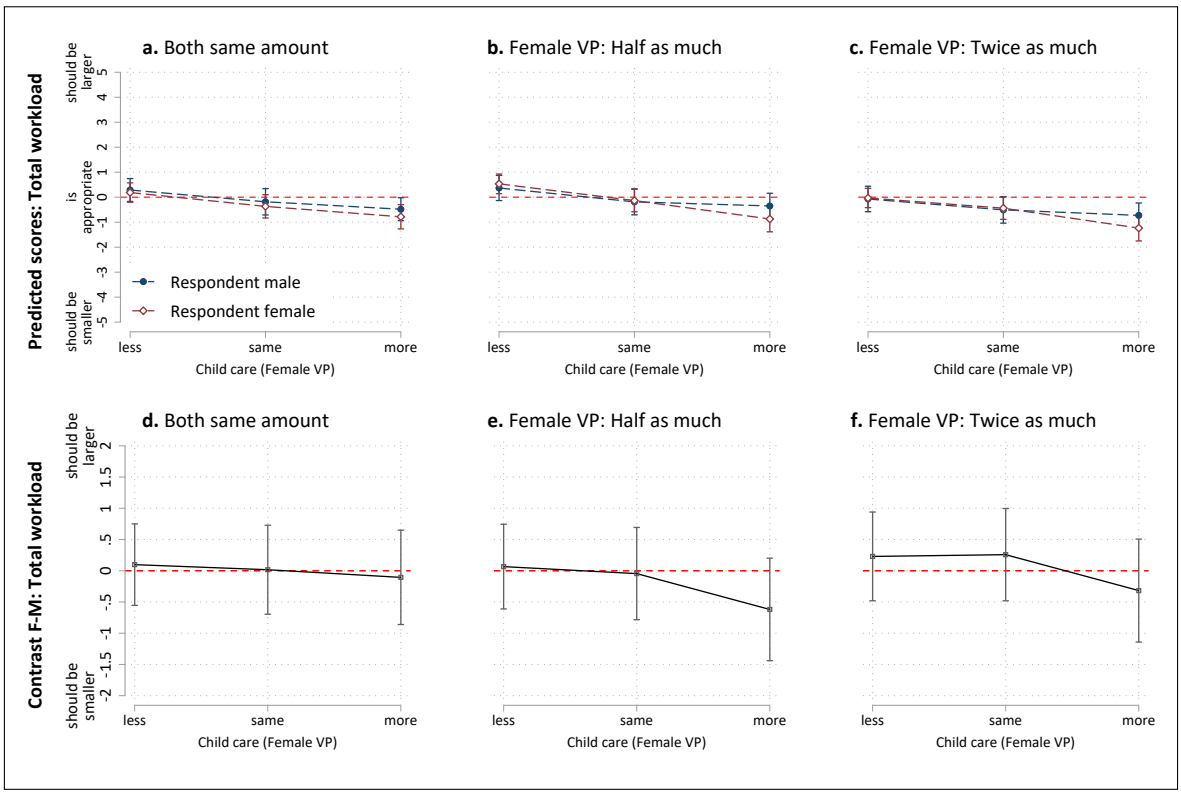


Figure 3.A2. Appropriateness of the Female VP's Total Workload by Share of Child Care and Relative Earnings

Note: Figure 3.A2 shows the appropriateness of the female VP's total workload by her share of child care (i.e., less, same or more than the male partner) and relative earnings. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's relative earnings: (a/d) both partners earn the same amount; (b/e) female VP earns half as much as her male partner; (c/f) female VP earns twice as much as her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.A2 (models 4a-c) in Appendix A.
Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

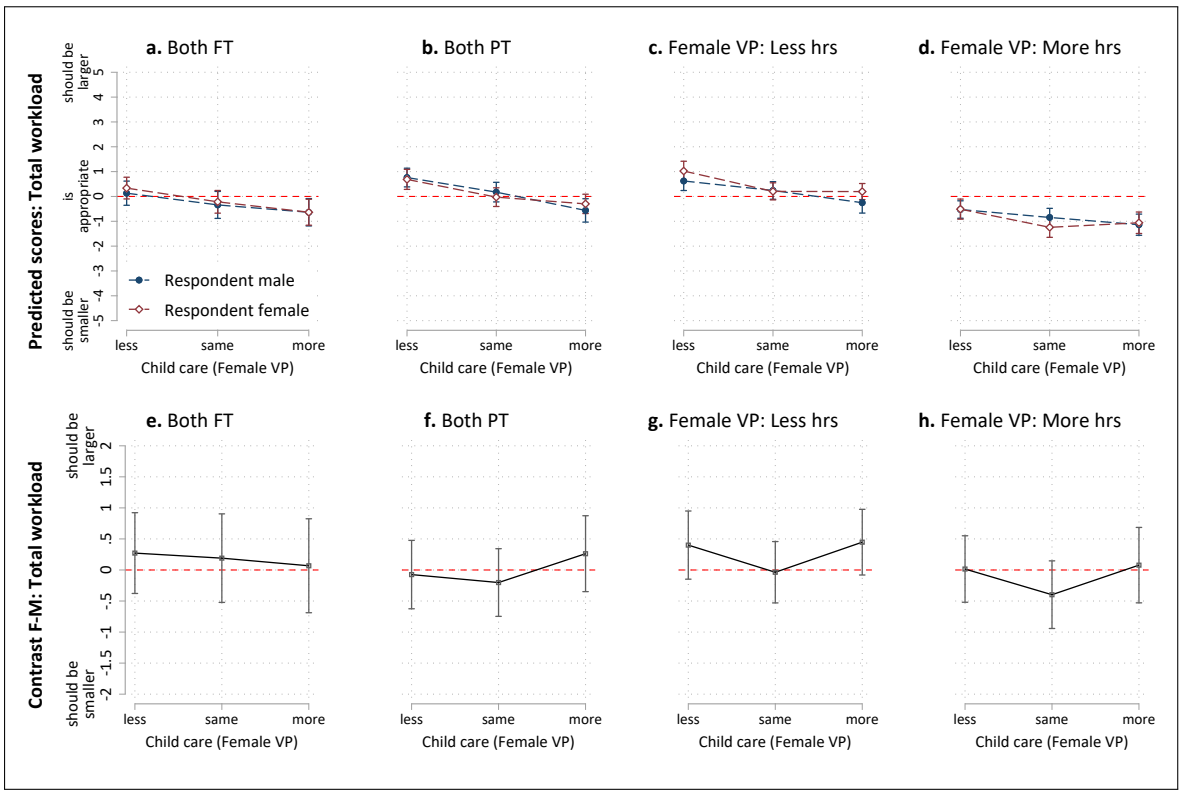


Figure 3.A3. Appropriateness of the Female VP's Total Workload by Share of Child Care and Paid Employment (Child = 8 Years Old)

Note: Figure 3.A3 shows the appropriateness of the female VP's total workload by share of child care (i.e., less, same, or more than the male partner) and paid employment. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's paid working hours: (a/e) both partners work full time (FT); (b/f) both partners work part time (PT); (c/g) the female VP has less paid working hours than her male partner; and (d/h) the female VP has more paid working hours than her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.A2 (models 4a-c) in Appendix A.
Number of vignette evaluations in total: 6,370, thereof 2,981 evaluations by men and 3,389 evaluations by women.

Table 3.A1. Descriptive Statistics

		Mean	s.d.	t-test
Gender role item question:				
Male respondents		4.149	0.867	$p<.001$
Female respondents		4.296	0.877	
Vignette variable: Appropriateness female VP's total workload:				
Male respondents		−0.268	1.939	$p=.401$
Female respondents		−0.311	2.084	
Descriptive statistics relating to vignette dimensions and levels:				
Share of child care (ref: both same)				
Female VP does 30% housework		0.201		
Female VP does 40% housework		0.199		
Female VP does 60% housework		0.193		
Female VP does 70% housework		0.206		
Share of housework (ref: both 50%)				
Female VP less child care		0.173		
Female VP more child care		0.163		
No information on child care		0.503		
Share of paid work (ref: both full time (FT))				
Both part time (PT) in paid work		0.220		
Female VP's paid work > Male VP's		0.331		
Female VP's paid work < Male VP's		0.340		
Relative earnings (ref: both same)				
Female VP half as much earnings		0.330		
Female VP twice as much earnings		0.330		
Marital status (ref: not married)				
Married		0.504		
Presence/Age of child (ref: no child)				
2-year-old child		0.331		
8-year-old child		0.329		
Characteristics of respondents:				
Male (ref: female)		0.468		
Age		36.142	8.020	
Married or in a relationship		0.771		
Number of individuals:	N			
Total	2,153			
Male respondents	1,008			
Female respondents	1,145			
Number of vignette evaluations:				
Total	6,370			
Male respondents	2,981			
Female respondents	3,389			

Table 3.A2. RI-Regression Model of Vignette Evaluations on Appropriateness of Female VP's Share of Total Workload

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Main effects: vignette factors						
Share of housework (ref: Female VP 50%)						
Female VP does 30%	1.137***	1.310***	1.136***	1.696***	1.901***	1.689***
	(0.092)	(0.090)	(0.092)	(0.271)	(0.274)	(0.271)
Female VP does 40%	0.592***	0.712***	0.592***	1.031***	0.792***	1.040***
	(0.084)	(0.087)	(0.084)	(0.245)	(0.237)	(0.247)
Female VP does 60%	−0.598***	−0.741***	−0.599***	−0.618**	−0.836***	−0.632**
	(0.082)	(0.085)	(0.082)	(0.237)	(0.213)	(0.238)
Female VP does 70%	−1.317***	−1.454***	−1.318***	−1.345***	−1.990***	−1.352***
	(0.094)	(0.091)	(0.094)	(0.301)	(0.250)	(0.300)
Share of child care (ref: both same)						
Female VP < Male VP	0.396***	0.743***	0.400***	0.488	0.543	0.472
	(0.094)	(0.095)	(0.094)	(0.337)	(0.300)	(0.338)
Female VP > Male VP	−0.421***	−0.290**	−0.417***	−0.281	−0.410	−0.296
	(0.104)	(0.098)	(0.104)	(0.349)	(0.341)	(0.350)
No info	−0.177	0.009	−0.176	0.139	−0.171	0.126
	(0.097)	(0.095)	(0.097)	(0.295)	(0.271)	(0.296)
Share of paid work (ref: both FT)						
Both PT	0.060	0.0680	0.059	0.568	0.181	0.563
	(0.105)	(0.102)	(0.105)	(0.300)	(0.259)	(0.301)
Female VP < Male VP	0.603***	0.793***	0.603***	0.636*	0.414	0.629*
	(0.102)	(0.094)	(0.103)	(0.295)	(0.261)	(0.296)
Female VP > Male VP	−0.861***	−0.879***	−0.861***	−0.442	−1.027***	−0.456
	(0.102)	(0.097)	(0.102)	(0.304)	(0.254)	(0.305)
Relative earnings (ref: same)						
Female VP half as much	0.0814	0.275***	0.0844	0.217	0.149	0.209
	(0.069)	(0.070)	(0.069)	(0.198)	(0.197)	(0.198)
Female VP twice as much	−0.413***	−0.250***	−0.412***	−0.252	−0.023	−0.264
	(0.070)	(0.070)	(0.070)	(0.195)	(0.209)	(0.195)
Marital status (ref: unmarried)						
Married	−0.061	0.029	−0.060	−0.048	0.020	−0.047
	(0.056)	(0.057)	(0.056)	(0.056)	(0.057)	(0.056)

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Presence/Age of child (ref: no child)						
2-year-old child	−0.033	−0.157	−0.036	−0.011	−0.185*	−0.015
	(0.093)	(0.089)	(0.093)	(0.094)	(0.089)	(0.094)
8-year-old child	−0.020	−0.066	−0.026	−0.013	−0.081	−0.020)
	(0.088)	(0.085)	(0.088)	(0.089)	(0.085)	(0.089)
Gender of VP (ref: man)						
Woman			−0.357			0.187
			(0.206)			(0.383)
Interactions:						
Female VP X						
Female VP 30% housework			0.175			0.204
			(0.129)			(0.384)
Female VP 40% housework			0.120			−0.269
			(0.121)			(0.343)
Female VP 60% housework			−0.139			−0.209
			(0.118)			(0.320)
Female VP 70% housework			−0.134			−0.640
			(0.131)			(0.391)
Female VP less child care			0.348**			0.081
			(0.133)			(0.451)
Female VP more child care			0.129			−0.123
			(0.143)			(0.489)
No information child care			0.190			−0.294
			(0.136)			(0.402)
Both PT paid work			0.006			−0.393
			(0.146)			(0.396)
Female VP more paid work			0.187			−0.227
			(0.139)			(0.395)
Female VP less paid work			−0.021			−0.589
			(0.141)			(0.397)
Female VP half as much earnings			0.189			−0.062
			(0.098)			(0.280)
Female VP twice as much earnings			0.162			0.241

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
			(0.099)			(0.285)
Married			0.088			0.066
			(0.080)			(0.080)
2-year-old child			−0.122			−0.170
			(0.129)			(0.129)
8-year-old child			−0.041			−0.062
			(0.122)			(0.123)
Both PT paid work X						
Female VP 30% housework				0.060	−0.450	0.067
				(0.314)	(0.301)	(0.315)
Female VP 40% housework				−0.366	−0.142	−0.369
				(0.283)	(0.259)	(0.284)
Female VP 60% housework				−0.026	−0.346	−0.012
				(0.258)	(0.251)	(0.258)
Female VP 70% housework				−0.739*	0.121	−0.719*
				(0.335)	(0.277)	(0.334)
Female VP less child care				0.110	0.170	0.115
				(0.358)	(0.333)	(0.359)
Female VP more child care				−0.425	0.134	−0.440
				(0.383)	(0.390)	(0.383)
No information child care				−0.477	0.013	−0.477
				(0.305)	(0.294)	(0.305)
F VP's paid work < M X						
Female VP 30% housework				−0.411	−0.628*	−0.407
				(0.296)	(0.285)	(0.296)
Female VP 40% housework				−0.164	0.004	−0.167
				(0.276)	(0.256)	(0.277)
Female VP 60% housework				0.401	0.143	0.413
				(0.255)	(0.229)	(0.256)
Female VP 70% housework				0.522	0.655*	0.530
				(0.317)	(0.258)	(0.316)
Female VP less child care				−0.099	0.272	−0.089
				(0.367)	(0.334)	(0.367)

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP more child care				−0.200	0.400	−0.196
				(0.363)	(0.358)	(0.364)
No information child care				−0.092	0.490	−0.088
				(0.300)	(0.281)	(0.301)
F VP's paid work > M						
Female VP 30% housework				−0.620*	−0.527	−0.619*
				(0.287)	(0.280)	(0.288)
Female VP 40% housework				−0.438	0.041	−0.443
				(0.267)	(0.255)	(0.268)
Female VP 60% housework				0.0002	0.172	0.016
				(0.272)	(0.256)	(0.273)
Female VP 70% housework				−0.033	0.476	−0.034
				(0.332)	(0.270)	(0.331)
Female VP less child care				−0.175	0.178	−0.154
				(0.361)	(0.315)	(0.361)
Female VP more child care				−0.008	0.587	0.004
				(0.378)	(0.366)	(0.379)
No information child care				−0.331	0.021	−0.317
				(0.304)	(0.268)	(0.305)
F VP half earnings X						
Female VP 30% housework				−0.520*	−0.006	−0.519*
				(0.215)	(0.209)	(0.215)
Female VP 40% housework				−0.175	0.132	−0.183
				(0.204)	(0.209)	(0.204)
Female VP 60% housework				−0.234	0.205	−0.233
				(0.195)	(0.204)	(0.195)
Female VP 70% housework				−0.126	0.106	−0.123
				(0.216)	(0.212)	(0.216)
Female VP less child care				0.071	0.116	0.084
				(0.219)	(0.224)	(0.219)
Female VP more child care				0.104	−0.321	0.131
				(0.238)	(0.226)	(0.238)
No information child care				0.103	0.156	0.113

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
				(0.190)	(0.182)	(0.190)
F VP twice earnings X						
Female VP 30% housework				−0.149	−0.272	−0.142
				(0.218)	(0.217)	(0.218)
Female VP 40% housework				−0.240	−0.332	−0.248
				(0.194)	(0.205)	(0.194)
Female VP 60% housework				−0.121	0.029	−0.120
				(0.195)	(0.198)	(0.195)
Female VP 70% housework				0.234	0.282	0.229
				(0.215)	(0.216)	(0.214)
Female VP less child care				−0.049	−0.140	−0.036
				(0.233)	(0.231)	(0.233)
Female VP more child care				0.052	−0.375	0.075
				(0.250)	(0.242)	(0.250)
No information child care				−0.200	−0.176	−0.184
				(0.198)	(0.195)	(0.198)
Female resp. X Both PT X						
Female VP 30% housework						−0.507
						(0.435)
Female VP 40% housework						0.245
						(0.385)
Female VP 60% housework						−0.328
						(0.360)
Female VP 70% housework						0.843
						(0.433)
Female VP less child care						0.047
						(0.490)
Female VP more child care						0.587
						(0.548)
No information child care						0.491
						(0.424)
Female resp. X F VP's paid work < M X						
Female VP 30% housework						−0.215

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
						(0.410)
Female VP 40% housework						0.190
						(0.378)
Female VP 60% housework						−0.266
						(0.344)
Female VP 70% housework						0.128
						(0.408)
Female VP less child care						0.356
						(0.497)
Female VP more child care						0.607
						(0.511)
No information child care						0.580
						(0.413)
Female resp. X F VP's paid work > M X						
Female VP 30% housework						0.102
						(0.401)
Female VP 40% housework						0.503
						(0.371)
Female VP 60% housework						0.164
						(0.374)
Female VP 70% housework						0.513
						(0.428)
Female VP less child care						0.332
						(0.480)
Female VP more child care						0.599
						(0.528)
No information child care						0.345
						(0.406)
Female resp. X F VP half earnings X						
Female VP 30% housework						0.517
						(0.300)
Female VP 40% housework						0.321
						(0.293)

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP 60% housework						0.442
						(0.283)
Female VP 70% housework						0.223
						(0.303)
Female VP less child care						0.031
						(0.313)
Female VP more child care						−0.451
						(0.329)
No information child care						0.042
						(0.263)
Female resp. X F VP twice earnings X						
Female VP 30% housework						−0.127
						(0.307)
Female VP 40% housework						−0.077
						(0.282)
Female VP 60% housework						0.154
						(0.278)
Female VP 70% housework						0.058
						(0.304)
Female VP less child care						−0.110
						(0.328)
Female VP more child care						−0.452)
						(0.347)
No information child care						0.004
						(0.278)
Constant	0.071	−0.286*	0.071	−0.338	−0.144	−0.321
	(0.155)	(0.136)	(0.155)	(0.294)	(0.244)	(0.295)
N (vignette evaluations)	2981	3389	6370	2981	3389	6370
N (individuals)	1008	1145	2153	1008	1145	2153
Wald test χ^2	817.9	1145.3	1969.9	1039.0	1392.7	2443.5
Wald p-value	0.000	0.000	0.000	0.000	0.000	0.000
Std. Dev. μ_j	0.599	0.720	0.666	0.580	0.721	0.659
Std. Dev. ϵ_{ij}	1.481	1.500	1.491	1.468	1.492	1.481

Continued on next page.

Table 3.A2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Interclass corr. ρ	0.141	0.187	0.166	0.135	0.189	0.165

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed $\chi^2(15)$ for columns 1 and 2, $\chi^2(31)$ for Column 3, $\chi^2(50)$ for columns 4a and 4b, and $\chi^2(101)$ for Column 4c.
ρ is the fraction of the variance of the unobserved component explained by the random effect.
Models (1) and (2) are separate models for male (1) and female (2) respondents.
Model (3) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables.
Model (4a) and (4b) are separate models for male (4a) and female (4b) respondents. They include two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care).
Model (4c) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables, two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care), and three-way interaction terms of a variable that indicates respondents' gender, the paid work vignette dimensions (i.e., share of paid work and relative earnings), and the unpaid work dimensions (i.e., shares of housework and child care).
Standard errors in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Appendix B: Validity and Robustness Checks

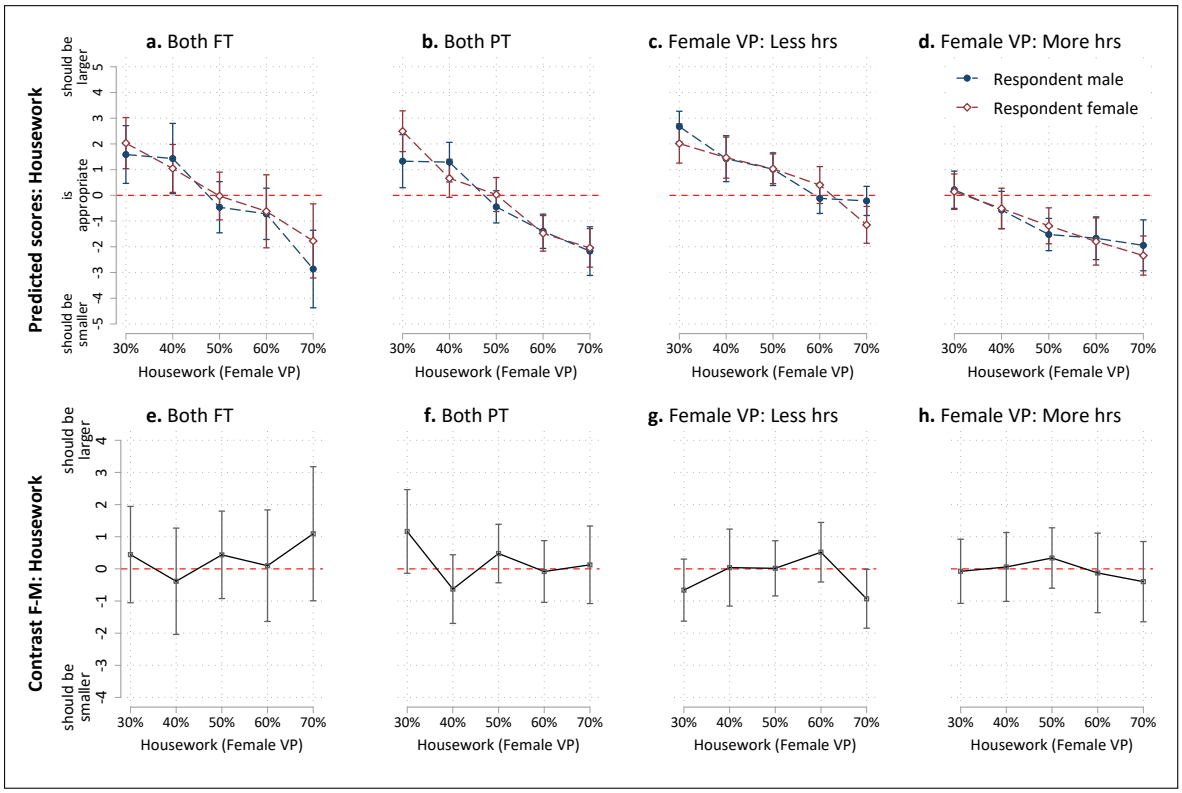


Figure 3.B1. Appropriateness of the Female VP's Housework by Housework Share and Paid Employment

Note: Figure 3.B1 shows the appropriateness of the female vignette person's (VP's) housework by housework share (i.e., ranging from 30% to 70% of total housework) and paid employment. The first row shows the predicted scores. The second row shows the contrast of female and male respondents. Each graph shows the results for a different relation of the VP's paid working hours: (a/e) both partners work full time (FT); (b/f) both partners work part time (PT); (c/g) the female VP has less paid working hours than her male partner; and (d/h) the female VP has more paid working hours than her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.B1 (models 4a-c) in Appendix B.
Number of vignette evaluations in total: 1,149, thereof 677 evaluations by men and 472 evaluations by women.

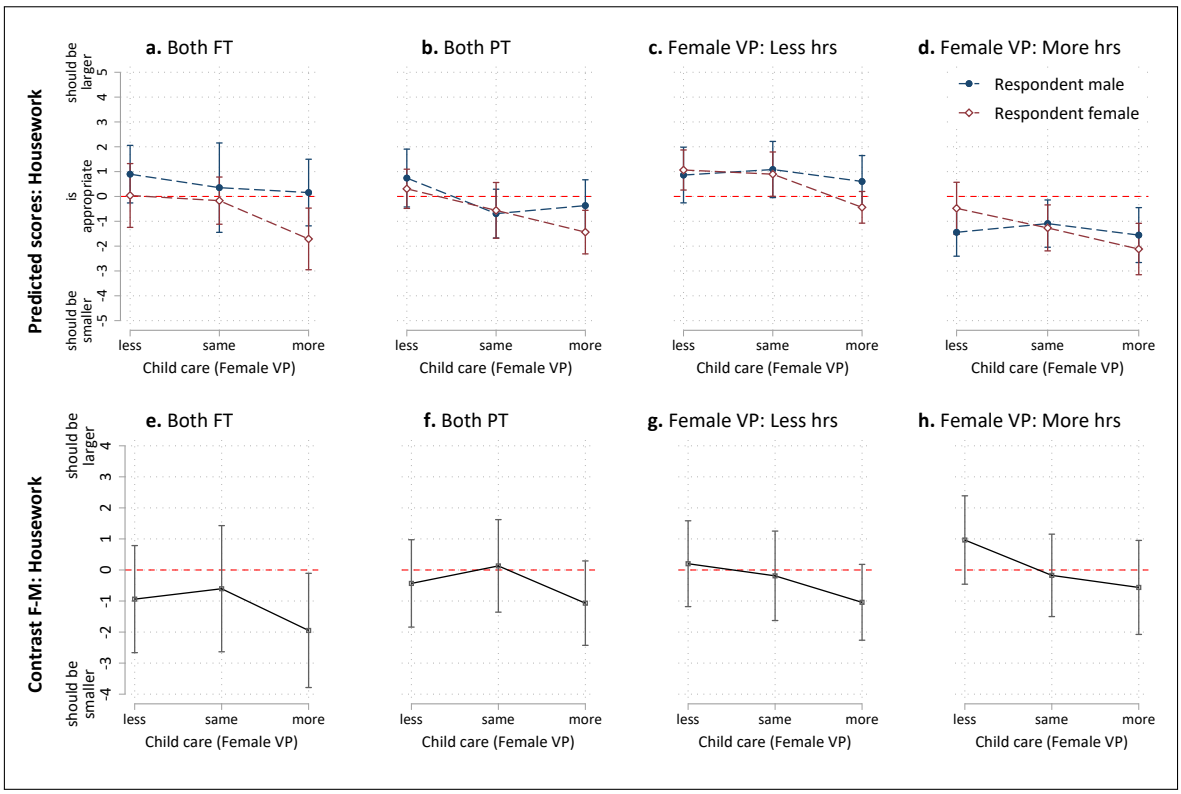


Figure 3.B2. Appropriateness of the Female VP's Housework by Share of Child Care and Paid Employment

Note: Figure 3.B2 shows the appropriateness of the female VP's housework by share of child care (i.e., less, same, or more than the male partner) and paid employment. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's paid working hours: (a/e) both partners work full time (FT); (b/f) both partners work part time (PT); (c/g) the female VP has less paid working hours than her male partner; and (d/h) the female VP has more paid working hours than her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.B1 (models 4a-c) in Appendix B.
Number of vignette evaluations in total: 1,149, thereof 677 evaluations by men and 472 evaluations by women.

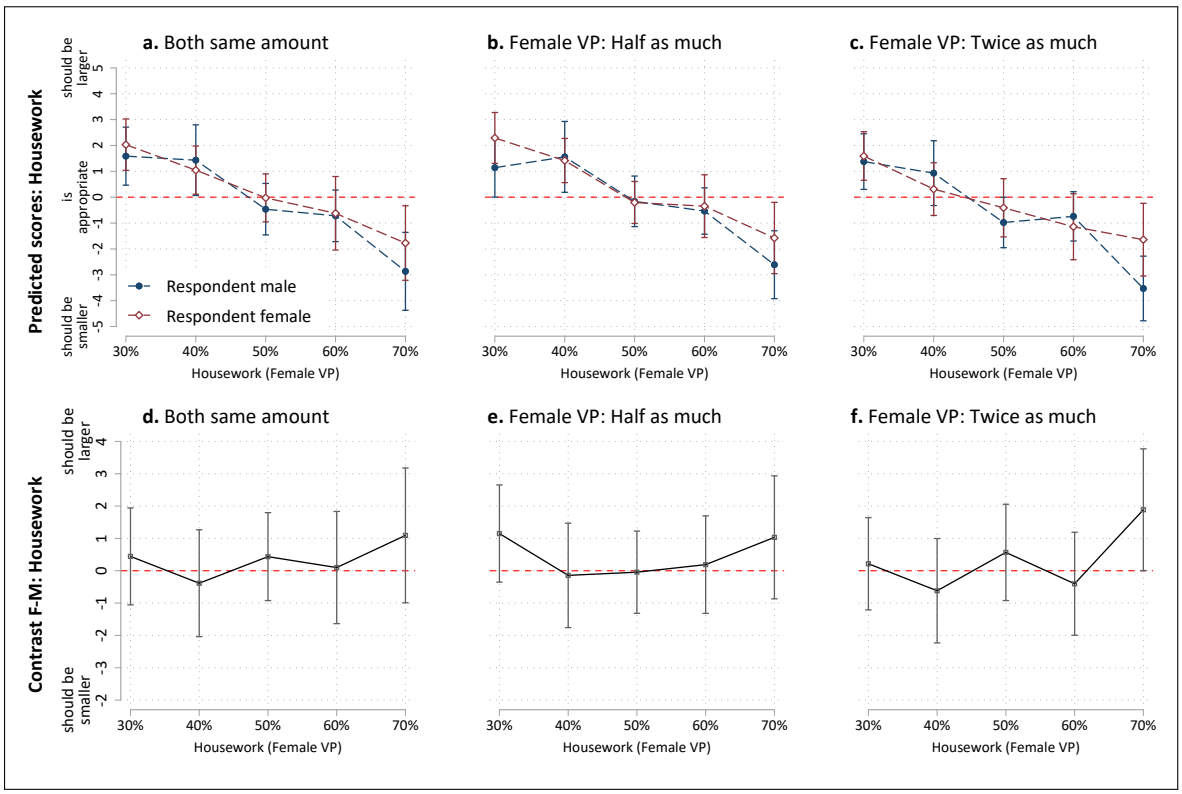


Figure 3.B3. Appropriateness of the Female VP's Housework by Housework Share and Relative Earnings

Note: Figure 3.B3 shows the appropriateness of the female VP's housework by housework share (i.e., ranging from 30% to 70% of total housework) and relative earnings. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's relative earnings: (a/d) both partners earn the same amount; (b/e) female VP earns half as much as her male partner; (c/f) female VP earns twice as much as her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.B1 (models 4a-c) in Appendix B.
Number of vignette evaluations in total: 1,149, thereof 677 evaluations by men and 472 evaluations by women.

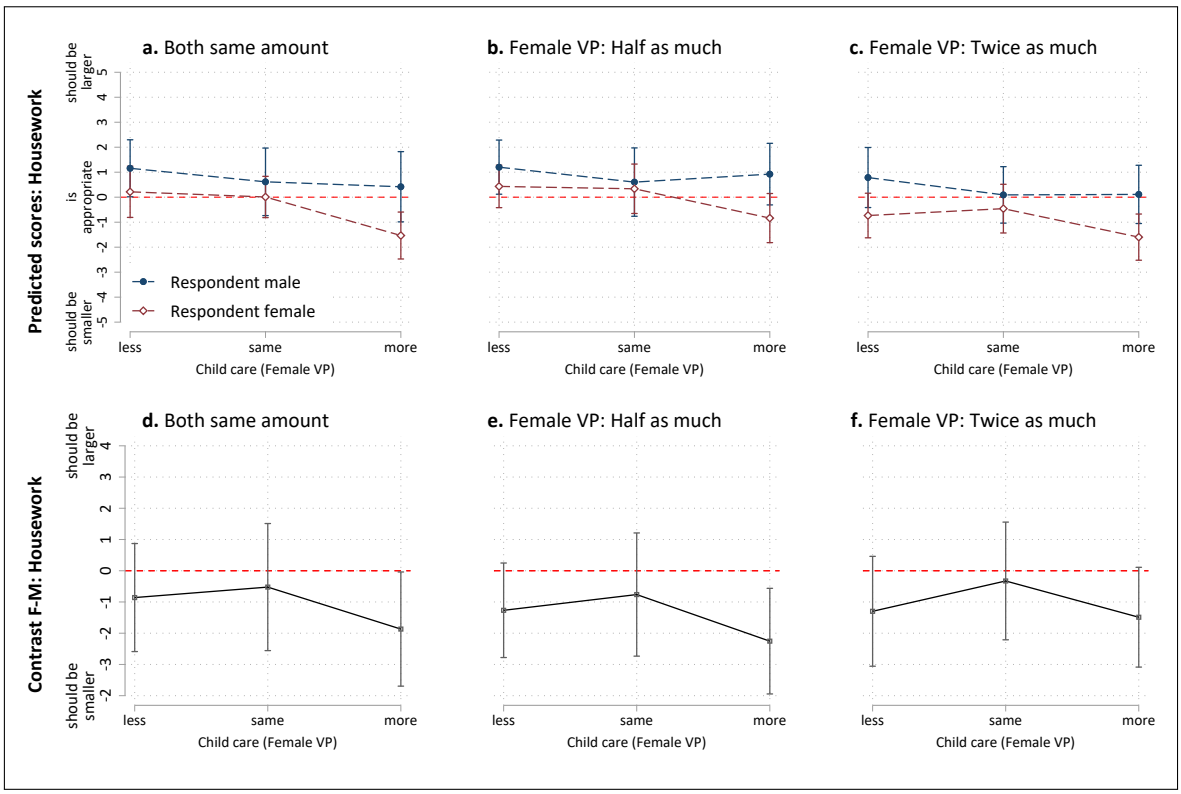


Figure 3.B4. Appropriateness of the Female VP's Housework by Share of Child Care and Relative Earnings

Note: Figure 3.B4 shows the appropriateness of the female VP's housework by her share of child care (i.e., less, same or more than the male partner) and relative earnings. The first row shows the predicted scores. The second row shows the contrast of female and male respondents.
Each graph shows the results for a different relation of the VP's relative earnings: (a/d) both partners earn the same amount; (b/e) female VP earns half as much as her male partner; (c/f) female VP earns twice as much as her male partner. The predictions were calculated based on separate regressions for male and female respondents. The contrast was estimated based on a joint RI model for male and female respondents. For the regression tables, see Table 3.B1 (models 4a-c) in Appendix B.
Number of vignette evaluations in total: 1,149, thereof 677 evaluations by men and 472 evaluations by women.

Table 3.B1. RI-Regression Model of Vignette Evaluations on Appropriateness of Female VP's Housework Share

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Main effects: vignette factors						
Share of housework (ref: Female VP 50%)						
Female VP does 30%	1.670***	1.560***	1.660***	2.043**	2.061***	2.050**
	(0.203)	(0.198)	(0.203)	(0.673)	(0.614)	(0.671)
Female VP does 40%	1.058***	0.727***	1.025***	1.820*	1.099*	1.895*
	(0.228)	(0.193)	(0.226)	(0.817)	(0.560)	(0.814)
Female VP does 60%	−0.639**	−0.717***	−0.656**	−0.305	−0.550	−0.257
	(0.206)	(0.208)	(0.204)	(0.677)	(0.760)	(0.679)
Female VP does 70%	−1.287***	−1.504***	−1.318***	−2.353**	−1.743*	−2.402**
	(0.241)	(0.179)	(0.236)	(0.873)	(0.724)	(0.862)
Share of child care (ref: both same)						
Female VP < Male VP	0.167	0.298	0.182	0.472	0.189	0.543
	(0.291)	(0.230)	(0.288)	(1.022)	(0.643)	(1.021)
Female VP > Male VP	−0.142	−0.850***	−0.113	−0.319	−1.520*	−0.197
	(0.273)	(0.220)	(0.267)	(1.098)	(0.600)	(1.075)
No info	−0.258	−0.248	−0.231	−1.124	−0.379	−1.046
	(0.244)	(0.234)	(0.238)	(0.787)	(0.531)	(0.787)
Share of paid work (ref: both FT)						
Both PT	−0.399	−0.147	−0.391	−1.270	−0.406	−1.146
	(0.298)	(0.231)	(0.304)	(0.927)	(0.535)	(0.919)
Female VP < Male VP	0.625*	0.689**	0.634*	0.523	1.051*	0.630
	(0.272)	(0.218)	(0.278)	(0.992)	(0.490)	(0.978)
Female VP > Male VP	−1.355***	−0.963***	−1.332***	−1.655	−1.096	−1.548
	(0.285)	(0.226)	(0.291)	(0.922)	(0.576)	(0.915)
Relative earnings (ref: same)						
Female VP half as much	0.134	0.300	0.121	0.223	−0.058	0.208
	(0.168)	(0.160)	(0.168)	(0.508)	(0.537)	(0.492)
Female VP twice as much	−0.378*	−0.444**	−0.387*	−0.615	−0.463	−0.662
	(0.176)	(0.147)	(0.175)	(0.547)	(0.459)	(0.538)
Marital status (ref: unmarried)						
Married	0.094	−0.0083	0.084	0.114	0.022	0.101
	(0.133)	(0.133)	(0.132)	(0.133)	(0.137)	(0.131)

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Presence/Age of child (ref: no child)						
2-year-old child	−0.186	−0.363	−0.172	−0.237	−0.534*	−0.229
	(0.230)	(0.232)	(0.229)	(0.239)	(0.243)	(0.236)
8-year-old child	−0.336	−0.117	−0.321	−0.446*	−0.244	−0.428*
	(0.193)	(0.213)	(0.198)	(0.200)	(0.217)	(0.200)
Gender of VP (ref: man)						
Woman			−0.124			−0.234
			(0.552)			(1.095)
Interactions:						
Female VP X						
Female VP 30% housework			−0.102			0.007
			(0.283)			(0.917)
Female VP 40% housework			−0.326			−0.822
			(0.298)			(0.992)
Female VP 60% housework			−0.072			−0.338
			(0.293)			(1.035)
Female VP 70% housework			−0.195			0.657
			(0.296)			(1.125)
Female VP less child care			0.122			−0.336
			(0.369)			(1.210)
Female VP more child care			−0.750*			−1.344
			(0.347)			(1.236)
No information child care			0.004			0.670
			(0.335)			(0.954)
Both PT paid work			0.254			0.735
			(0.382)			(1.068)
Female VP more paid work			0.064			0.415
			(0.353)			(1.096)
Female VP less paid work			0.388			0.430
			(0.369)			(1.085)
Female VP half as much earnings			0.194			−0.239
			(0.231)			(0.730)
Female VP twice as much earnings			−0.056			0.196

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
			(0.229)			(0.710)
Married			−0.092			−0.081
			(0.187)			(0.190)
2-year-old child			−0.167			−0.289
			(0.326)			(0.340)
8-year-old child			0.228			0.198
			(0.291)			(0.297)
Both PT paid work X						
Female VP 30% housework				−0.237	0.410	−0.274
				(0.734)	(0.627)	(0.732)
Female VP 40% housework				−0.033	−0.437	−0.156
				(0.797)	(0.639)	(0.790)
Female VP 60% housework				−0.684	−0.928	−0.692
				(0.689)	(0.789)	(0.688)
Female VP 70% housework				0.668	−0.340	0.680
				(0.877)	(0.775)	(0.862)
Female VP less child care				0.939	0.671	0.889
				(0.952)	(0.738)	(0.954)
Female VP more child care				0.642	0.648	0.521
				(1.053)	(0.763)	(1.026)
No information child care				1.263	0.455	1.162
				(0.794)	(0.589)	(0.789)
F VP's paid work < Male V X						
Female VP 30% housework				−0.355	−1.075	−0.391
				(0.688)	(0.600)	(0.689)
Female VP 40% housework				−1.390	−0.620	−1.485
				(0.862)	(0.674)	(0.857)
Female VP 60% housework				−0.780	−0.058	−0.877
				(0.720)	(0.729)	(0.721)
Female VP 70% housework				1.167	−0.418	1.169
				(0.840)	(0.769)	(0.827)
Female VP less child care				−0.715	−0.054	−0.766
				(0.988)	(0.677)	(0.984)

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP more child care				−0.192	0.194	−0.284
				(1.119)	(0.704)	(1.093)
No information child care				0.909	−0.008	0.849
				(0.800)	(0.658)	(0.793)
F VP's paid work > M X						
Female VP 30% housework				−0.285	−0.746	−0.309
				(0.668)	(0.644)	(0.669)
Female VP 40% housework				−0.821	−0.425	−0.942
				(0.813)	(0.677)	(0.813)
Female VP 60% housework				0.201	−0.049	0.112
				(0.715)	(0.783)	(0.714)
Female VP 70% housework				1.943*	0.582	1.981*
				(0.928)	(0.775)	(0.919)
Female VP less child care				−0.855	0.577	−0.893
				(0.993)	(0.736)	(0.993)
Female VP more child care				−0.189	0.664	−0.265
				(1.087)	(0.682)	(1.068)
No information child care				0.543	−0.064	0.487
				(0.760)	(0.569)	(0.760)
F VP half earnings X						
Female VP 30% housework				−0.770	0.444	−0.748
				(0.528)	(0.434)	(0.520)
Female VP 40% housework				−0.188	0.534	−0.175
				(0.536)	(0.442)	(0.527)
Female VP 60% housework				−0.122	0.434	−0.117
				(0.465)	(0.498)	(0.454)
Female VP 70% housework				−0.070	0.382	−0.047
				(0.535)	(0.429)	(0.517)
Female VP less child care				0.086	−0.073	0.056
				(0.541)	(0.611)	(0.533)
Female VP more child care				0.537	0.377	0.517
				(0.550)	(0.621)	(0.542)
No information child care				0.086	−0.123	0.095

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
				(0.410)	(0.548)	(0.409)
F VP twice earnings X						
Female VP 30% housework				0.280	−0.056	0.303
				(0.502)	(0.484)	(0.498)
Female VP 40% housework				0.004	−0.369	0.013
				(0.503)	(0.496)	(0.498)
Female VP 60% housework				0.473	−0.166	0.494
				(0.457)	(0.506)	(0.456)
Female VP 70% housework				−0.172	0.508	−0.150
				(0.498)	(0.472)	(0.485)
Female VP less child care				0.152	−0.466	0.152
				(0.601)	(0.573)	(0.594)
Female VP more child care				0.193	0.412	0.218
				(0.646)	(0.531)	(0.643)
No information child care				0.111	0.091	0.148
				(0.453)	(0.507)	(0.450)
Female resp. X Both PT X						
Female VP 30% housework						0.679
						(0.971)
Female VP 40% housework						−0.284
						(1.024)
Female VP 60% housework						−0.221
						(1.062)
Female VP 70% housework						−1.008
						(1.160)
Female VP less child care						−0.229
						(1.211)
Female VP more child care						0.144
						(1.283)
No information child care						−0.694
						(0.992)
Female resp. X F VP's paid work < M X						
Female VP 30% housework						−0.685

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
						(0.919)
Female VP 40% housework						0.845
						(1.099)
Female VP 60% housework						0.839
						(1.038)
Female VP 70% housework						−1.604
						(1.130)
Female VP less child care						0.726
						(1.200)
Female VP more child care						0.491
						(1.304)
No information child care						−0.835
						(1.038)
Female resp. X F VP's paid work > M X						
Female VP 30% housework						−0.421
						(0.935)
Female VP 40% housework						0.543
						(1.065)
Female VP 60% housework						−0.126
						(1.075)
Female VP 70% housework						−1.392
						(1.203)
Female VP less child care						1.475
						(1.241)
Female VP more child care						0.956
						(1.274)
No information child care						−0.528
						(0.957)
Female resp. X F VP half earnings X						
Female VP 30% housework						1.187
						(0.677)
Female VP 40% housework						0.724
						(0.689)

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP 60% housework						0.571
						(0.677)
Female VP 70% housework						0.420
						(0.672)
Female VP less child care						−0.169
						(0.813)
Female VP more child care						−0.147
						(0.827)
No information child care						−0.242
						(0.685)
Female resp. X F VP twice earnings X						
Female VP 30% housework						−0.358
						(0.699)
Female VP 40% housework						−0.363
						(0.709)
Female VP 60% housework						−0.631
						(0.688)
Female VP 70% housework						0.664
						(0.679)
Female VP less child care						−0.635
						(0.830)
Female VP more child care						0.184
						(0.836)
No information child care						−0.065
						(0.682)
Constant	0.273	0.168	0.260	0.677	0.354	0.583
	(0.414)	(0.364)	(0.413)	(0.968)	(0.514)	(0.962)
N (vignette evaluations)	472	677	1149	472	667	1149
N (individuals)	158	226	384	158	226	384
Wald test χ^2	259.9	364.2	622.3	612.0	519.3	1113.1
Wald p-value	0.000	0.000	0.000	0.000	0.000	0.000
Std. Dev. μ_j	0.246	0.629	0.511	0.341	0.638	0.547
Std. Dev. ϵ_{ij}	1.483	1.591	1.548	1.450	1.581	1.530

Continued on next page.

Table 3.B1 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Interclass corr. ρ	0.027	0.135	0.098	0.052	0.140	0.113

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed $\chi^2(15)$ for columns 1 and 2, $\chi^2(31)$ for Column 3, $\chi^2(50)$ for columns 4a and 4b, and $\chi^2(101)$ for Column 4c.
ρ is the fraction of the variance of the unobserved component explained by the random effect.
Models (1) and (2) are separate models for male (1) and female (2) respondents.
Model (3) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables.
Model (4a) and (4b) are separate models for male (4a) and female (4b) respondents. They include two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care).
Model (4c) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables, two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care), and three-way interaction terms of a variable that indicates respondents' gender, the paid work vignette dimensions (i.e., share of paid work and relative earnings), and the unpaid work dimensions (i.e., shares of housework and child care).
Standard errors in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Table 3.B2. OLS-Regression Model of Vignette Evaluations on Appropriateness of Female VP's Share of Total Workload

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Main effects: vignette factors						
Share of housework (ref: Female VP 50%)						
Female VP does 30%	1.141***	1.329***	1.141***	1.731***	1.832***	1.731***
	(0.093)	(0.093)	(0.093)	(0.273)	(0.265)	(0.272)
Female VP does 40%	0.589***	0.709***	0.589***	0.981***	0.600*	0.981***
	(0.086)	(0.093)	(0.086)	(0.239)	(0.250)	(0.239)
Female VP does 60%	−0.593***	−0.712***	−0.593***	−0.543*	−0.888***	−0.543*
	(0.083)	(0.089)	(0.083)	(0.235)	(0.225)	(0.234)
Female VP does 70%	−1.310***	−1.442***	−1.310***	−1.307***	−2.007***	−1.307***
	(0.095)	(0.094)	(0.095)	(0.311)	(0.251)	(0.311)
Share of child care (ref: both same)						
Female VP < Male VP	0.374***	0.786***	0.374***	0.569	0.637*	0.569
	(0.098)	(0.095)	(0.098)	(0.339)	(0.306)	(0.339)
Female VP > Male VP	−0.448***	−0.264**	−0.448***	−0.206	−0.492	−0.206
	(0.104)	(0.100)	(0.104)	(0.347)	(0.357)	(0.347)
No info	−0.186	0.054	−0.186	0.204	−0.137	0.204
	(0.099)	(0.095)	(0.099)	(0.293)	(0.284)	(0.293)
Share of paid work (ref: both FT)						
Both PT	0.062	0.041	0.062	0.598*	0.073	0.598*
	(0.105)	(0.104)	(0.105)	(0.301)	(0.261)	(0.301)
Female VP < Male VP	0.606***	0.753***	0.606***	0.679*	0.302	0.679***
	(0.102)	(0.095)	(0.102)	(0.293)	(0.267)	(0.293)
Female VP > Male VP	−0.857***	−0.917***	−0.857***	−0.370	−1.189***	−0.370
	(0.104)	(0.100)	(0.104)	(0.304)	(0.267)	(0.304)
Relative earnings (ref: same)						
Female VP half as much	0.062	0.262***	0.062	0.256	0.129	0.256
	(0.071)	(0.072)	(0.071)	(0.201)	(0.206)	(0.201)
Female VP twice as much	−0.419***	−0.249***	−0.419***	−0.187	−0.028	−0.187
	(0.071)	(0.072)	(0.071)	(0.199)	(0.208)	(0.199)
Marital status (ref: unmarried)						
Married	−0.063	0.020	−0.063	−0.052	0.013	−0.052
	(0.057)	(0.058)	(0.057)	(0.057)	(0.058)	(0.057)

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Presence/Age of child (ref: no child)						
2-year-old child	−0.017	−0.162	−0.017	0.008	−0.189*	0.008
	(0.095)	(0.093)	(0.095)	(0.096)	(0.092)	(0.096)
8-year-old child	0.014	−0.074	0.014	0.019	−0.090	0.019
	(0.089)	(0.087)	(0.089)	(0.090)	(0.087)	(0.090)
Gender of VP (ref: man)						
Woman			−0.360			0.361
			(0.211)			(0.387)
Interactions:						
Female VP X						
Female VP 30% housework			0.188			0.102
			(0.131)			(0.380)
Female VP 40% housework			0.120			−0.381
			(0.126)			(0.346)
Female VP 60% housework			−0.119			−0.345
			(0.122)			(0.325)
Female VP 70% housework			−0.132			−0.700
			(0.133)			(0.399)
Female VP less child care			0.412**			0.069
			(0.137)			(0.457)
Female VP more child care			0.183			−0.286
			(0.145)			(0.498)
No information child care			0.241			−0.342
			(0.137)			(0.408)
Both PT paid work			−0.021			−0.525
			(0.148)			(0.398)
Female VP more paid work			0.147			−0.377
			(0.139)			(0.397)
Female VP less paid work			−0.061			−0.820*
			(0.144)			(0.404)
Female VP half as much earnings			0.200*			−0.127
			(0.101)			(0.288)
Female VP twice as much earnings			0.170			0.159

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
			(0.101)			(0.287)
Married			0.083			0.066
			(0.082)			(0.081)
2-year-old child			−0.145			−0.196
			(0.133)			(0.133)
8-year-old child			−0.088			−0.109
			(0.125)			(0.126)
Both PT paid work X						
Female VP 30% housework				0.025	−0.359	0.025
				(0.313)	(0.294)	(0.313)
Female VP 40% housework				−0.347	0.022	−0.347
				(0.283)	(0.271)	(0.282)
Female VP 60% housework				−0.103	−0.285	−0.103
				(0.260)	(0.259)	(0.260)
Female VP 70% housework				−0.843*	0.158	−0.843*
				(0.347)	(0.281)	(0.346)
Female VP less child care				0.087	0.097	0.087
				(0.361)	(0.339)	(0.360)
Female VP more child care				−0.346	0.261	−0.346
				(0.382)	(0.401)	(0.381)
No information child care				−0.476	0.027	−0.476
				(0.305)	(0.300)	(0.304)
F VP's paid work < M X						
Female VP 30% housework				−0.434	−0.579*	−0.434
				−(0.298)	(0.283)	(0.298)
Female VP 40% housework				−0.154	0.178	−0.154
				(0.272)	(0.269)	(0.271)
Female VP 60% housework				0.335	0.187	0.335
				(0.255)	(0.241)	(0.255)
Female VP 70% housework				0.477	0.693**	0.477
				(0.325)	(0.263)	(0.325)
Female VP less child care				−0.149	0.225	−0.149
				(0.370)	(0.344)	(0.369)

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP more child care				−0.217	0.498	−0.217
				(0.365)	(0.371)	(0.365)
No information child care				−0.112	0.503	−0.112
				(0.300)	(0.293)	(0.300)
F paid work > M X						
Female VP 30% housework				−0.627*	−0.440	−0.627*
				(0.287)	(0.279)	(0.287)
Female VP 40% housework				−0.414	0.210	−0.414
				(0.262)	(0.271)	(0.261)
Female VP 60% housework				−0.081	0.249	−0.081
				(0.277)	(0.268)	(0.277)
Female VP 70% housework				−0.032	0.509	−0.032
				(0.341)	(0.280)	(0.340)
Female VP less child care				−0.281	0.184	−0.281
				(0.364)	(0.325)	(0.363)
Female VP more child care				−0.067	0.726	−0.067
				(0.379)	(0.382)	(0.379)
No information child care				−0.399	0.084	−0.399
				(0.302)	(0.280)	(0.302)
F VP half earnings X						
Female VP 30% housework				−0.530*	0.041	−0.530*
				(0.222)	(0.219)	(0.222)
Female VP 40% housework				−0.130	0.185	−0.130
				(0.207)	(0.225)	(0.207)
Female VP 60% housework				−0.236	0.239	−0.236
				(0.199)	(0.214)	(0.198)
Female VP 70% housework				−0.142	0.048	−0.142
				(0.221)	(0.217)	(0.221)
Female VP less child care				0.005	0.100	0.005
				(0.224)	(0.229)	(0.224)
Female VP more child care				−0.038	−0.311	−0.038
				(0.238)	(0.235)	(0.238)
No information child care				0.051	0.146	0.051

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
				(0.191)	(0.185)	(0.191)
F VP twice earnings X						
Female VP 30% housework				−0.186	−0.241	−0.186
				(0.222)	(0.215)	(0.221)
Female VP 40% housework				−0.199	−0.264	−0.199
				(0.197)	(0.210)	(0.197)
Female VP 60% housework				−0.125	0.075	−0.125
				(0.199)	(0.203)	(0.199)
Female VP 70% housework				0.260	0.332	0.260
				(0.220)	(0.221)	(0.219)
Female VP less child care				−0.119	−0.193	−0.119
				(0.236)	(0.235)	(0.236)
Female VP more child care				−0.067	−0.399	−0.067
				(0.253)	(0.245)	(0.253)
No information child care				−0.285	−0.213	−0.285
				(0.200)	(0.198)	(0.200)
Female resp. X Both PT X						
Female VP 30% housework						−0.384
						(0.430)
Female VP 40% housework						0.369
						(0.391)
Female VP 60% housework						−0.183
						(0.367)
Female VP 70% housework						1.001*
						(0.446)
Female VP less child care						0.010
						(0.495)
Female VP more child care						0.606
						(0.553)
No information child care						0.504
						(0.427)
Female resp. X F VP's paid work < M X						
Female VP 30% housework						−0.145

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
						(0.411)
Female VP 40% housework						0.332
						(0.382)
Female VP 60% housework						−0.148
						(0.351)
Female VP 70% housework						0.216
						(0.418)
Female VP less child care						0.374
						(0.505)
Female VP more child care						0.715
						(0.520)
No information child care						0.614
						(0.419)
Female resp. X Fe VP's paid work > M X						
Female VP 30% housework						0.187
						(0.401)
Female VP 40% housework						0.624
						(0.376)
Female VP 60% housework						0.330
						(0.386)
Female VP 70% housework						0.541
						(0.441)
Female VP less child care						0.465
						(0.487)
Female VP more child care						0.792
						(0.538)
No information child care						0.484
						(0.412)
Female resp. X F VP half earnings X						
Female VP 30% housework						0.571
						(0.312)
Female VP 40% housework						0.315
						(0.305)

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP 60% housework						0.476
						(0.292)
Female VP 70% housework						0.190
						(0.309)
Female VP less child care						0.095
						(0.320)
Female VP more child care						−0.272
						(0.335)
No information child care						0.094
						(0.266)
Female resp. X F VP twice earnings X						
Female VP 30% housework						−0.056
						(0.309)
Female VP 40% housework						−0.065
						(0.288)
Female VP 60% housework						0.200
						(0.285)
Female VP 70% housework						0.072
						(0.311)
Female VP less child care						−0.074
						(0.333)
Female VP more child care						−0.333
						(0.352)
No information child care						0.072
						(0.282)
Constant	0.073	−0.287*	0.073	−0.420	−0.060	−0.420
	(0.160)	(0.138)	(0.160)	(0.292)	(0.254)	(0.292)
N (vignette evaluations)	2981	3389	6370	2981	3389	6370
N (individuals)	1008	1145	2153	1008	1145	2153
F-test	51.57	75.28	61.42	19.95	27.82	23.67
F-test p-value	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.329	0.368	0.351	0.350	0.380	0.367

Continued on next page.

Table 3.B2 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.

Note: The F test relates to the joint significance of all coefficients in the column. It is distributed F(15, 1007) for columns 1, F(15, 1144) for Column 2, F(31, 2152) for column 3, F(50, 1007) for column 4a, F(50, 1144) for column 4b, and F(101, 2152) for column 4c.
Models (1) and (2) are separate models for male (1) and female (2) respondents.
Model (3) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables.
Model (4a) and (4b) are separate models for male (4a) and female (4b) respondents. They include two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care).
Model (4c) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables, two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care), and three-way interaction terms of a variable that indicates respondents' gender, the paid work vignette dimensions (i.e., share of paid work and relative earnings), and the unpaid work dimensions (i.e., shares of housework and child care).
Standard errors in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

Table 3.B3. RI-Regression Model of Vignette Evaluations on Appropriateness of Female VP's Share of Total Workload (First Vignette Not Included.)

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Main effects: vignette factors						
Share of housework (ref: Female VP 50%)						
Female VP does 30%	1.211***	1.325***	1.215***	1.599***	2.030***	1.593***
	(0.109)	(0.108)	(0.109)	(0.296)	(0.369)	(0.295)
Female VP does 40%	0.643***	0.680***	0.650***	0.937***	0.949***	0.937***
	(0.102)	(0.107)	(0.102)	(0.240)	(0.254)	(0.240)
Female VP does 60%	−0.599***	−0.780***	−0.597***	−0.621*	−0.584*	−0.650*
	(0.099)	(0.101)	(0.099)	(0.263)	(0.266)	(0.266)
Female VP does 70%	−1.343***	−1.420***	−1.342***	−1.582***	−1.863***	−1.605***
	(0.114)	(0.107)	(0.114)	(0.352)	(0.322)	(0.351)
Share of child care (ref: both same)						
Female VP < Male VP	0.427***	0.693***	0.430***	0.605	0.598	0.595
	(0.116)	(0.121)	(0.116)	(0.345)	(0.361)	(0.346)
Female VP > Male VP	−0.465***	−0.275*	−0.464***	−0.317	−0.325	−0.341
	(0.124)	(0.119)	(0.124)	(0.356)	(0.415)	(0.358)
No info	−0.142	−0.097	−0.144	0.144	−0.212	0.137
	(0.119)	(0.118)	(0.120)	(0.299)	(0.318)	(0.300)
Share of paid work (ref: both FT)						
Both PT	0.149	0.018	0.151	0.251	0.354	0.249
	(0.124)	(0.123)	(0.125)	(0.303)	(0.312)	(0.303)
Female VP < Male VP	0.625***	0.748***	0.627***	0.540	0.621	0.540
	(0.119)	(0.117)	(0.119)	(0.302)	(0.318)	(0.303)
Female VP > Male VP	−0.769***	−0.894***	−0.764***	−0.469	−0.842**	−0.486
	(0.120)	(0.118)	(0.120)	(0.324)	(0.323)	(0.324)
Relative earnings (ref: same)						
Female VP half as much	0.196*	0.302***	0.197*	0.376	0.236	0.361
	(0.084)	(0.085)	(0.084)	(0.236)	(0.242)	(0.237)
Female VP twice as much	−0.364***	−0.225**	−0.363***	−0.112	−0.025	−0.137
	(0.086)	(0.084)	(0.086)	(0.262)	(0.243)	(0.263)
Marital status (ref: unmarried)						
Married	−0.012	0.044	−0.011	−0.002	0.040	−0.001

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)
Presence/Age of child (ref: no child)						
2-year-old child	0.023	−0.317**	0.022	0.028	−0.340**	0.026
	(0.114)	(0.110)	(0.114)	(0.117)	(0.109)	(0.117)
8-year-old child	0.064	−0.217*	0.055	0.077	−0.340**	0.067
	(0.110)	(0.110)	(0.110)	(0.111)	(0.109)	(0.111)
Gender of VP (ref: man)						
Woman			0.006			0.220
			(0.250)			(0.431)
Interactions:						
Female VP X						
Female VP 30% housework			0.110			0.429
			(0.153)			(0.470)
Female VP 40% housework			0.033			−0.012
			(0.148)			(0.351)
Female VP 60% housework			−0.175			0.048
			(0.141)			(0.378)
Female VP 70% housework			−0.075			−0.254
			(0.157)			(0.477)
Female VP less child care			0.267			−0.003
			(0.168)			(0.502)
Female VP more child care			0.195			−0.011
			(0.172)			(0.549)
No information child care			0.052			−0.354
			(0.168)			(0.439)
Both PT paid work			−0.134			0.082
			(0.175)			(0.435)
Female VP more paid work			0.117			0.060
			(0.167)			(0.440)
Female VP less paid work			−0.133			−0.385
			(0.169)			(0.459)
Female VP half as much earnings			0.101			−0.132
			(0.120)			(0.339)

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP twice as much earnings			0.138			0.101
			(0.120)			(0.358)
Married			0.055			0.042
			(0.098)			(0.097)
2-year-old child			−0.341*			−0.370*
			(0.159)			(0.161)
8-year-old child			−0.272			−0.309*
			(0.156)			(0.156)
Both PT paid work X						
Female VP 30% housework				0.275	−0.528	0.282
				(0.330)	(0.381)	(0.331)
Female VP 40% housework				−0.067	−0.340	−0.058
				(0.293)	(0.301)	(0.294)
Female VP 60% housework				0.238	−0.733*	0.266
				(0.289)	(0.306)	(0.293)
Female VP 70% housework				−0.437	−0.270	−0.421
				(0.398)	(0.357)	(0.398)
Female VP less child care				0.191	0.073	0.182
				(0.384)	(0.388)	(0.384)
Female VP more child care				−0.150	0.266	−0.145
				(0.406)	(0.466)	(0.408)
No information child care				−0.263	−0.020	−0.277
				(0.313)	(0.335)	(0.314)
F VP's paid work < M X						
Female VP 30% housework				−0.220	−0.692	−0.222
				(0.318)	(0.372)	(0.318)
Female VP 40% housework				−0.073	−0.005	−0.071
				(0.278)	(0.296)	(0.278)
Female VP 60% housework				0.414	−0.231	0.437
				(0.279)	(0.276)	(0.282)
Female VP 70% housework				0.821*	0.499	0.841*
				(0.384)	(0.328)	(0.384)
Female VP less child care				−0.172	0.040	−0.170

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
				(0.386)	(0.384)	(0.386)
Female VP more child care				−0.134	0.443	−0.134
				(0.386)	(0.443)	(0.389)
No information child care				−0.119	0.307	−0.126
				(0.301)	(0.330)	(0.301)
F VP's paid work > M X						
Female VP 30% housework				−0.641*	−0.574	−0.624
				(0.323)	(0.367)	(0.323)
Female VP 40% housework				−0.569*	−0.086	−0.560*
				(0.282)	(0.302)	(0.283)
Female VP 60% housework				−0.0009	−0.182	0.036
				(0.312)	(0.312)	(0.314)
Female VP 70% housework				0.225	0.227	0.236
				(0.402)	(0.352)	(0.402)
Female VP less child care				−0.257	−0.093	−0.241
				(0.386)	(0.377)	(0.386)
Female VP more child care				0.035	0.710	0.065
				(0.399)	(0.456)	(0.402)
No information child care				−0.147	−0.007	−0.142
				(0.324)	(0.321)	(0.324)
F VP half earnings X						
Female VP 30% housework				−0.331	−0.147	−0.325
				(0.249)	(0.255)	(0.248)
Female VP 40% housework				−0.046	−0.023	−0.050
				(0.246)	(0.250)	(0.246)
Female VP 60% housework				−0.182	0.224	−0.174
				(0.234)	(0.247)	(0.234)
Female VP 70% housework				−0.152	0.264	−0.131
				(0.264)	(0.250)	(0.264)
Female VP less child care				−0.082	0.298	−0.069
				(0.254)	(0.283)	(0.255)
Female VP more child care				−0.128	−0.606*	−0.112
				(0.278)	(0.285)	(0.279)

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
No information child care				−0.030	0.114	−0.021
				(0.223)	(0.231)	(0.224)
F VP twice earnings X						
Female VP 30% housework				−0.146	−0.320	−0.136
				(0.278)	(0.268)	(0.278)
Female VP 40% housework				−0.078	−0.481	−0.069
				(0.264)	(0.250)	(0.265)
Female VP 60% housework				−0.282	0.110	−0.275
				(0.247)	(0.239)	(0.247)
Female VP 70% housework				0.141	0.480	0.154
				(0.273)	(0.265)	(0.273)
Female VP less child care				−0.092	−0.067	−0.078
				(0.298)	(0.296)	(0.298)
Female VP more child care				−0.071	−0.529	−0.048
				(0.315)	(0.286)	(0.315)
No information child care				−0.307	−0.115	−0.287
				(0.251)	(0.234)	(0.252)
Female resp. X Both PT X						
Female VP 30% housework						−0.807
						(0.503)
Female VP 40% housework						−0.258
						(0.422)
Female VP 60% housework						−0.975*
						(0.423)
Female VP 70% housework						0.159
						(0.536)
Female VP less child care						−0.095
						(0.548)
Female VP more child care						0.444
						(0.619)
No information child care						0.263
						(0.460)
Female resp. X F VP's paid work < M X						

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Female VP 30% housework						−0.473
						(0.488)
Female VP 40% housework						0.080
						(0.408)
Female VP 60% housework						−0.648
						(0.395)
Female VP 70% housework						−0.345
						(0.505)
Female VP less child care						0.224
						(0.547)
Female VP more child care						0.603
						(0.589)
No information child care						0.443
						(0.448)
Female resp. X F VP's paid work > M X						
Female VP 30% housework						0.047
						(0.487)
Female VP 40% housework						0.493
						(0.416)
Female VP 60% housework						−0.183
						(0.444)
Female VP 70% housework						−0.005
						(0.535)
Female VP less child care						0.176
						(0.541)
Female VP more child care						0.681
						(0.608)
No information child care						0.148
						(0.457)
Female resp. X F VP half earnings X						
Female VP 30% housework						0.198
						(0.356)
Female VP 40% housework						0.043

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
						(0.351)
Female VP 60% housework						0.402
						(0.341)
Female VP 70% housework						0.378
						(0.364)
Female VP less child care						0.358
						(0.382)
Female VP more child care						−0.494
						(0.399)
No information child care						0.138
						(0.322)
Female resp. X F VP twice earnings X						
Female VP 30% housework						−0.174
						(0.386)
Female VP 40% housework						−0.394
						(0.364)
Female VP 60% housework						0.390
						(0.344)
Female VP 70% housework						0.332
						(0.380)
Female VP less child care						0.0002
						(0.421)
Female VP more child care						−0.473
						(0.425)
No information child care						0.177
						(0.344)
Constant	−0.065	−0.061	−0.068	−0.341	−0.122	−0.321
	(0.182)	(0.171)	(0.183)	(0.305)	(0.302)	(0.307)
N (vignette evaluations)	1985	2251	4236	1985	2251	4236
N (individuals)	1002	1137	2139	1002	1137	2139
Wald test χ^2	722.4	851.7	1587.9	926.7	1059.0	1993.3
Wald p-value	0.000	0.000	0.000	0.000	0.000	0.000
Std. Dev. μ_j	0.609	0.788	0.710	0.610	0.790	0.710

Continued on next page.

Table 3.B3 – *Continued from previous page.*

	(1) Men	(2) Women	(3) Diff.	(4a) Men interac.	(4b) Women interact.	(4c) Diff. interact.
Std. Dev. ϵ_{ij}	1.452	1.468	1.460	1.439	1.449	1.445
Interclass corr. ρ	0.150	0.224	0.191	0.151	0.229	0.195

Note: The Wald test relates to the joint significance of all coefficients in the column. It is distributed $\chi^2(15)$ for columns 1 and 2, $\chi^2(31)$ for Column 3, $\chi^2(50)$ for columns 4a and 4b, and $\chi^2(101)$ for Column 4c.
ρ is the fraction of the variance of the unobserved component explained by the random effect.
Models (1) and (2) are separate models for male (1) and female (2) respondents.
Model (3) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables.
Model (4a) and (4b) are separate models for male (4a) and female (4b) respondents. They include two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care).
Model (4c) is a joint model for male and female respondents. It includes two-way interaction terms of a variable that indicates respondents' gender with the relevant vignette variables, two-way interaction terms of the paid work vignette dimensions (i.e., share of paid work and relative earnings) with the unpaid work dimensions (i.e., shares of housework and child care), and three-way interaction terms of a variable that indicates respondents' gender, the paid work vignette dimensions (i.e., share of paid work and relative earnings), and the unpaid work dimensions (i.e., shares of housework and child care).
Standard errors in parentheses. * $p < .05$; ** $p < .01$; *** $p < .001$

4

Different Order, Different Results? The Effects of Dimension Order in Factorial Survey Experiments.

(Sabine Düval and Thomas Hinz)

Accepted manuscript of an article published in:
Field Methods 2020, Vol. 21(1):23-37.
https://doi.org/10.1177/1525822X19886827

4 Different Order, Different Results? The Effects of Dimension Order in Factorial Survey Experiments.

Abstract Factorial surveys are widely used in the social sciences to measure respondents' attitudes, beliefs, or behavioral intentions. In such surveys, respondents evaluate short descriptions of hypothetical situations, persons, or objects that vary across several dimensions. An important prerequisite of the method's validity is that respondents are able to deal with the highly complex task created by the need to consider several variable dimensions within one coherent judgment. We analyze the effects of the order in which dimensions are presented in running text vignettes. An experimental setup with four order treatments was randomly allocated to 787 respondents (based on a random sample of register data), yielding 3,119 vignette evaluations. The analyses compare respondent groups across age, education, and response speed. Overall, there is no strong evidence for order effects. However, we find a slight tendency for fast responders to be more prone to recency effects.

4.1 Introduction

Factorial surveys have been widely used in the social sciences to measure respondents' attitudes, beliefs, or decisions. In such surveys, respondents are asked about one or several vignettes (scenarios) describing hypothetical situations, objects, or persons. According to the experimental design, across the vignettes, attributes (dimensions) vary in their values (levels). This allows the estimation of the impact of each dimension's value on the respondents' judgments. The method is used in many disciplines, such as sociology, political science, criminology, and consumer research. While the method is widespread (for a review, see Wallander, 2009), research on methodological issues still lags behind (Auspurg and Jäckle, 2017). The evaluation task, based on a multifactorial design, is relatively complex, and this could call into question substantive conclusions drawn from such studies.

We focus on one prominent issue in survey research, namely order effects, and analyze whether the order in which the vignette dimensions are presented affects results. So far, there have been only two studies on this using samples consisting of university students (Auspurg and Jäckle, 2017; Robbins and Kiser, 2018). We use a more heterogeneous respondent sample, which allows the study of possible variations according to respondents' cognitive abilities (measured by proxies of age and degree of education). Respondents were asked to evaluate a factorial survey module on the substantive issue of potential financial support of university students. In four experimental splits, we varied the order of the seven different vignette dimensions. In general, we found no evidence of order effects. This was true for several subgroups of respondents (higher vs. lower cognitive ability, different speeds of response). A closer look revealed a slight tendency toward recency effects (i.e., vignette dimensions presented at the very end of a vignette having a comparably higher impact). To document our analytical strategy, we provide comprehensive Appendices, to which we refer interested readers.

4.2 Overview of Current Research

4.2.1 Order Effects in Survey Research

The cognitive revolution in survey research of the 1980s and 1990s (e.g., Beniger and Gusek, 1995) raised awareness of the fact that responsive behavior is likely to be affected by the cognitive burden of taking a survey. In general, order effects arise when the order of response alternatives affects results. In complex question formats, survey questions demand the completion of different cognitive tasks: Respondents need to remember the question and the different response alternatives, evaluate the consequences of each alternative, and choose the one they most agree with (e.g., Schwarz, 2007; Schwarz et al., 1991). Given complex tasks, respondents might be unable to actively remember

all presented alternatives at the same time (Ayidiya and McClendon, 1990). While alternatives presented first in a list enter long-term memory, alternatives presented at the end are more likely to enter short-term memory. When there is no delay between the presentation of the alternatives and the answering of the question, information from the short-term memory (i.e., alternatives presented at the end of a list) is more easily recalled than information from the longterm memory (Schwarz et al., 1991). This leads to recency effects. Primacy effects are the opposite of recency effects.

Heterogeneity of respondents might moderate order effects. Recalling information should be easier for individuals who are more motivated, who have more experience solving complex cognitive problems, or who even enjoy solving them (Krosnick, 1989). Also, respondents with greater knowledge have more skills available to help them choose the optimal response (McClendon, 1991). At the same time, people with less working memory capacity are expected to be more prone to order effects (Knäuper, 1999).

Most studies have measured working memory capacity by using proxies like education or age (e.g., Holbrook et al., 2007; Krosnick and Alwin, 1987). As there is a strong correlation between education and scores on tests of cognitive skills, education should be a good proxy of cognitive skills (Ceci, 1991). Also, age-related reductions in cognitive resources starting around the age of 60 make it more difficult to hold large amounts of information in short-term memory (Schaie, 1996; Schwarz and Knäuper, 2000, p. 248).

Respondents who are cognitively challenged when answering survey questions might follow a satisficing strategy, instead of providing an optimal answer, to decrease their cognitive effort (Krosnick, 1991). Instead of thinking about all the different options simultaneously and trying to arrive at a single answer based on all of them, respondents might adopt the first reasonable alternative and base their answer disproportionately on this item. Such a satisficing approach takes much less memory capacity and therefore can be detected through lower response times.

4.2.2 Order Effects in Factorial Surveys

In factorial survey experiments, respondents read the vignettes, which are displayed in a paper questionnaire or on a screen in web surveys, in a self-administered mode. Factorial survey experiments are considered cognitively challenging because respondents have to take a number of (varying) dimensions into account before making a judgment. To our knowledge, there have been only two methodological articles that have focused on the order of vignette dimensions (Auspurg and Jäckle, 2017; Robbins and Kiser, 2018). Auspurg and Jäckle (2017) find some indications of order effects but only when the survey task becomes more complex (in terms of dimensions and number of vignettes to evaluate). Moreover, they do not find any signs of primacy or recency effects. Robbins and Kiser (2018) do not find any evidence of order effects in a similar experimental setting.

Both studies demonstrate this with vignettes in a tabular format (rarely used in the standard applications of factorial surveys where running text dominates) and with data from homogeneous samples of university students and hence samples with high levels of cognitive capacity. We expand the methodological research on order effects to the presentation of running text vignettes and to a general population sample with respondents aged between 18 and 83 years and with different levels of education. This allows us to examine to what extent possible order effects are moderated by respondents' cognitive ability.

As indicated above, it is worth considering one specific order effect, namely whether the last vignette attribute evokes recency effects. When answering a question after reading a vignette description spanning several questionnaire lines, information in a person's short-term memory might be much more easily recalled, which would favor such recency effects.

When it comes to respondents' moderators, we expect older respondents, and those with less formal education, to be more prone to order effects than younger respondents or respondents with a higher degree of education. Finally, we expect fast responders to be more prone to order effects because of satisficing strategies.

4.3 Data and Methods

4.3.1 Data and Sample

To answer our questions, we rely on a general population sample. The *Konstanzer Bürgerbefragung* is an annual panel survey of the citizens of the German city of Konstanz that has been conducted since 2008 (Hinz et al., 2012), with different main survey topics such as neighborhood quality, life satisfaction, and political participation. We use data collected in 2011 via computer-assisted self-interview, when the thematic focus was the educational system. All respondents were randomly drawn from the population register, with an overall return rate of 37.1%.

The data collection included a factorial survey to evaluate opinions on who should be supported by state-financed scholarships when applying for university admission. Each respondent was asked to evaluate four vignettes describing hypothetical applicants. The total number of observations was 3,640 vignette evaluations by 910 respondents. We excluded five respondents with missing values on the vignette evaluation. Furthermore, following standard procedures, only plausible response time measures were used: All time values above the 99th percentile and under the 1st percentile were dropped; in a second step, all values were excluded that were two standard deviations above or under the mean (i.e., 80 respondents; Mayerl et al., 2005). Additionally, 38 respondents with missing values on the proxy measures of cognitive ability (age and educational level) were dropped. Excluded respondents did not differ with regard to

education and response time from nonexcluded participants (there was a small difference in age—excluded respondents were 1.63 years younger on average). Finally, we use an analysis sample of 787 respondents, who evaluated 3,119 vignettes.

4.3.2 Factorial Survey Module

All hypothetical situations describe students at the University of Konstanz applying for scholarships by the state to support the students with €500 per month (for an example vignette, see Figure 4.A1 in Appendix A). To ensure that the respondents differentiated between the students, they were told that only a third of the applicants could be supported, and it was up to the respondents to decide whether the award of a scholarship was reasonable or not. The question asked at the end of each vignette was as follows: "In your opinion, should this person rightfully receive a scholarship or not?" using an 11-point rating scale ranging from -5 *(not at all)* to +5 *(by all means)*.

The students who applied for scholarships were characterized by seven variable dimensions that are known to influence allocation decisions in education: (1) the name of the student (clearly indicating both gender and ethnic background), (2) place of birth, (3) average grade at high school graduation (grade point average), (4) parents' income, (5) number of siblings, (6) students' work experience, and (7) career orientation. Every dimension varied on two or three levels (e.g., parents' income was either low, middle, or high; for all dimension levels, see Table 4.A1 in Appendix A).

Of the set of 2,916 possible vignette scenarios, we generated a sample of 200 vignettes using a D-efficient sampling technique that reflects both orthogonality and level balance (Auspurg and Hinz, 2015; Kuhfeld et al., 1994, for more information on D-efficiency, see Appendix E). With a D-efficiency value of 97.7, our design has enough statistical power for our research. We split the sample of 200 vignettes into 50 questionnaire versions (sets) of four vignettes, again using a D-efficient sampling technique that maximizes orthogonality and level balance.

To mirror standard practices in factorial survey research (see Wallander, 2009), we decided to present vignettes in running text (and not in a tabular order). Thus, a complete randomization of the order of all dimensions was not feasible while keeping a logical sentence structure in a running text. For instance, introducing a vignette person requires information on their names at the beginning of the text, followed by information that would be expected (e.g., place of birth and school achievement). It would be odd to start running texts describing students with dimensions of career orientations or the number of siblings and providing names and place of birth at the end of the text. Of 720 possible arrangements of dimension orders (keeping names at the beginning), we decided on four selected, different orders of dimensions as treatment groups (see Table 4.1). The four order treatments each represent smooth vignette texts, given the structure of the German language, and enable us to test for possible overall

order effects and order effects across respondent groups. Each of the 50 sets of vignettes was assigned at least once to all four treatment groups. Additionally, the order of the vignettes within each set was randomized for each respondent. The respondents were randomly allocated to one of the sets and treatment groups.

For each of the four treatments (dimension order), we have at least 709 vignette evaluations from at least 180 respondents (for detailed information on the sample size, see Table 4.A2 in Appendix A). Respondents' characteristics, such as age, educational level, response time, gender, and migration background, did not systematically vary across the four treatment groups.

Table 4.1. Information on Treatment Split

Position of dimension	Treatment split: Four order groups			
	1	2	3	4
1	Name	Name	Name	Name
2	Place of birth	GPA	GPA	Place of birth
3	GPA	Place of birth	Place of birth	GPA
4	Parents' income	Parents' income	Work experience	Work experience
5	Siblings	Siblings	Career orientation	Parents' income
6	Work experience	Career orientation	Parents' income	Siblings
7	Career orientation	Work experience	Siblings	Career orientation

Note: GPA= grade point average

4.3.3 Operationalization of Respondents' Cognitive Ability and Response Time

We measured cognitive ability via the proxies of respondents' age and education level. Following Auspurg and Hinz (2015), respondents who are 60 years and older are classified as one subgroup, while the reference is under 60 years. As Narayan and Krosnick (1996) found that respondents with low education levels (those who did not complete high school) showed stronger order effects than respondents with a medium or high level of education (who did not differ from each other), we decided to split respondents accordingly: Respondents who at least graduated from high school are said to have a high education level, and all others are in the reference group of lower education level. Further analyses revealed that the results are robust to a different criterion (e.g., academic degree). In all estimations of age effects, education is controlled and vice versa.

The composition of the subgroups of age and education shows that the mean age of the analysis sample is 45.98 years (*s.d.* = 4.97); 2,478 vignettes were evaluated by respondents under the age of 60 years, while older respondents evaluated 641 vignettes. More than three-fourths (75.22%) of the sample at least graduated from high school, while 24.78% only graduated from junior high school or below (for more information on the sample, see Table 4.A2 in Appendix A).

To measure response time, we used time stamps to calculate the time span respondents needed to read and evaluate the vignettes. When analyzing response time, it must be taken into account that the baseline speed for answering survey questions differs between respondents because of cognitive factors but also because of technical aspects relevant to an online survey, such as differing Internet speed or device effects (Mayerl et al., 2005). Only if the raw time needed for answering the four vignettes is calibrated by each respondent's individual baseline speed can a meaningful comparison of response time across different respondents be made. In short, this measure reveals whether the respondents needed a particularly short or long time to answer the vignettes, in comparison with the time they needed to answer a different part of the survey (Sauer et al., 2011). In this study, baseline speed is estimated through four questions on life satisfaction that represent a medium cognitive burden. This baseline speed is used to adjust the raw time span by using a residual index approach (Urban and Mayerl, 2007). Higher values on the residual index indicate that respondents needed more time to answer the vignettes than their baseline speed would indicate; accordingly, lower values indicate a faster response time. By means of a median split, the respondents were classified into the two subgroups fast and slow responders.

4.3.4 Analytical Strategy

In all estimations, vignette evaluations are regressed on all seven vignette dimensions by using ordinary least squares (OLS) regressions. To adjust for the unequal variances of error terms caused by the clustering of evaluations within respondents, we use clustered standard errors (Rogers, 1994). Detailed information on all models and further robustness checks can be found in extensive tables in the Appendices B-E.

We test for overall order effects within the full sample and within each of the six subgroups (under 60 years, 60 years and older, higher education level, lower education level, fast responders, and slow responders). All tests are estimated for the full sample and for each of the six subgroups, separately. In a first step, we estimate OLS regressions for all four order treatments. Differences in regression coefficients between the four order splits would give us first indications of order effects. Second, we estimate a joint model of all four order splits, including two-way interaction terms of each dimension with an indicator variable for the order split. Here, a joint Wald test (also known as a "Chow test," see Wooldridge, 2009) allows us to test whether the interaction terms

are all jointly zero, meaning there are no statistically significant differences between the order splits caused by the order of the dimensions. Third, Wald tests for each individual vignette dimension allow an even closer look at dimension order effects: Now we compare each individual dimension's coefficient across all four treatment groups. Fourth, with "moderator tests" (Auspurg and Jäckle, 2017), we test whether possible differences in the extent of order effects between the relevant subgroups are statistically significant. The test logic is comparable to that described above, but instead of two-way interaction terms of the relevant dimension and the order indicator, we introduce three-way interaction terms, with the respective moderator variable (age, education, or speed) as an additional variable.

To focus on the occurrence of recency effects, we apply a similar strategy, again separately for the full sample and for all subgroups. For each of the three dimensions that are placed on the last position of the vignette in at least one order split (i.e., career orientation, work experience, and siblings), we again estimate separate OLS regressions including a two-way interaction term of the relevant dimension with an indicator as to whether the dimension was in the last position. A statistically significant interaction term indicates a recency effect. Again, we use joint Wald tests to test whether the interaction terms are zero.

4.4 Results

4.4.1 Is There an Overall Order Effect?

For the full sample, Figure 4.1 depicts the influence of each dimension on the vignette evaluation for all four order splits. With the exception of the dimension level "one sibling," there is no clearly visible difference between the four treatment groups. The result of the joint Wald test supports the conclusion that there is no overall dimension order effect in the full sample ($F = 0.86$, $p = .697$; see Table 4.2).

Table 4.2. Wald Tests and Moderator Tests for Full Sample and Subgroups

	Wald Test		Moderator Test	
	N	*F (p)*	*N*	*F(p)*
Full sample	2,030	0.86 (.697)		
Under 60 years	1,615	1.30 (.124)	2,030	1.17 (.235)
60 years and older	415	0.80 (.774)		
Higher education	1,515	1.06 (.380)	2,030	0.82 (.755)
Lower education	515	0.72 (.868)		
Fast responders	1,014	1.31 (.123)	2,030	0.85 (.704)
Slow responders	1,016	0.61 (.956)		

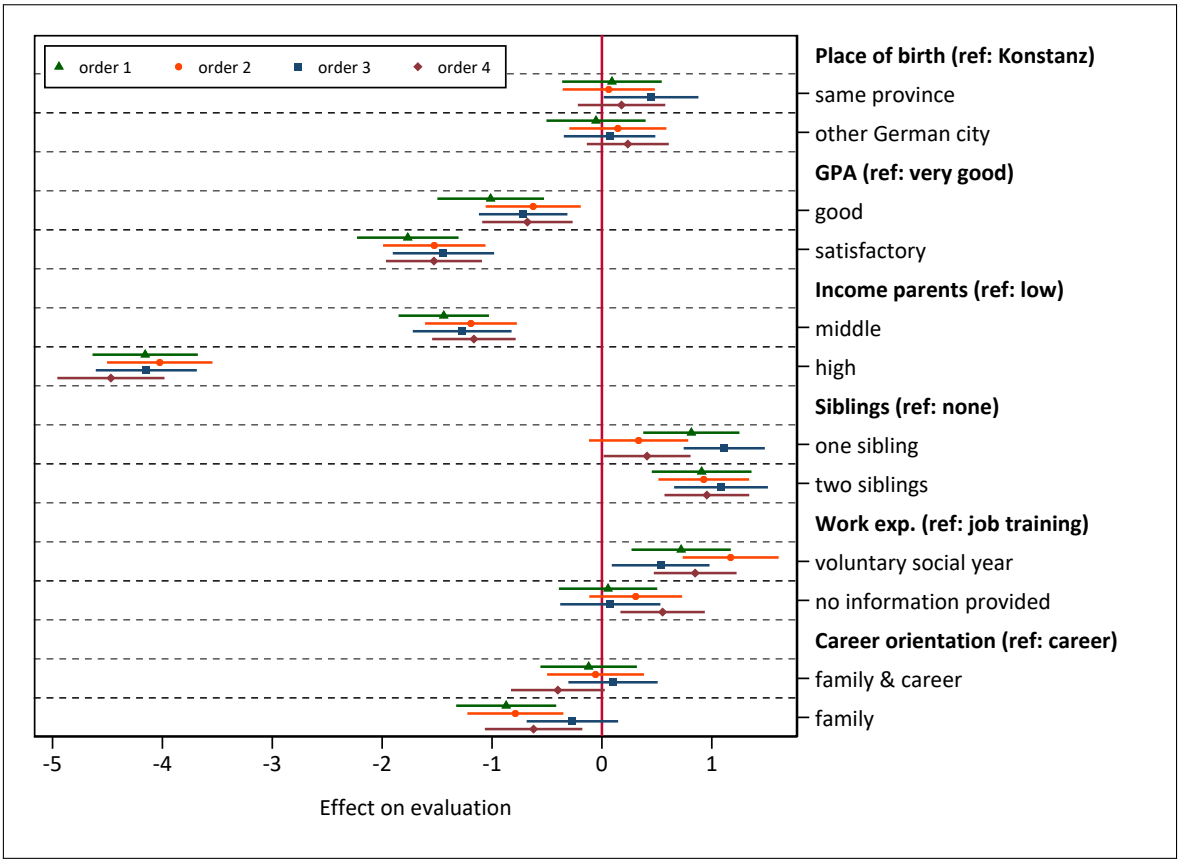


Figure 4.1. Coefficient Plot of Regression of Vignette Evaluation by Treatment Splits

Note: The figure shows regression coefficients of the vignette dimensions by the four treatment splits (see Table 4.B1 in Appendix B for the regression tables). Only dimensions whose positions change between treatment splits are shown. Standard errors with 95% confidence intervals are attached.

In Table 4.3, Wald tests for each individual vignette dimension show that only 3 of all 66 possible comparisons of the dimensions' positions indicate statistically significant order effects. Those 4.5% statistically significant interaction terms are expected to be statistically significant by chance only. However, it is worth noting that all three effects indicate recency effects (i.e., their impact on the vignette evaluations is higher if the dimension is the last one posed in the vignette).

Within the six respondent subgroups, the results of the OLS regressions do not point toward order effects (see Tables 4.B2 - 4.B7 in Appendix B for the regression tables). Moreover, Wald tests reveal no statistically significant order effects (see Table 4.2). Also, additional Wald tests for each interaction term individually do not point toward subgroups that are particularly prone to dimension order effects (not displayed, see Tables 4.C2 - 4.C7 in Appendix C). The only exception are fast responders, where 5 of 66 possible dimensions show statistically significant order effects. Those 7.6% are slightly more than could be expected to occur by chance (for a detailed interpretation, see Appendix E).

The moderator tests do not support the idea that respondents with lower cognitive

Table 4.3. Individual Wald Tests (Full Sample)

Comparison	Dimension	χ^2 (p)
Order 2 versus 3	Siblings (ref.: no siblings)	
	One sibling	6.987 (.008)
	Work experience (ref.: job training)	
	Voluntary social year	4.175 (.041)
Order 3 versus 4	Siblings (ref.: no siblings)	
	One sibling	4.349 (.037)

Note: We compare the dimensions' coefficients between two order groups. For the full table with all dimension combinations, see Table 4.C1 in Appendix C and for a more detailed interpretation of the results, see Appendix E.

abilities are more prone to order effects compared with respondents with higher cognitive abilities (see Table 4.2). In addition, the two speed groups do not differ in their susceptibility to order effects.

4.4.2 Is There a Recency Effect?

In Figure 4.2 (section A), the point estimates for the full sample suggest that the dimension siblings shows statistically significant recency effects for one of the dimension levels: The effect size increases when the dimension level "one sibling" is in the last position ($p = .012$). The point estimates for work experience (section B) and career orientation (section C) do not show any statistically significant recency effects. All results are supported by joint Wald tests (see Table 4.D1 in Appendix D).

The results in Figure 4.2 indicate recency effects in the dimension work experience ($p = .021$; section B) for people over 60, while younger people show recency effects in one level of the dimension siblings ($p = .009$; section A). Concerning education, if any effect is visible, it is that more educated respondents show more recency effects (for a detailed interpretation, see Appendix E). In sum, the results do not clearly suggest that respondents with lower cognitive abilities are consistently more prone to recency effects than respondents with higher cognitive abilities. Concerning response time, *fast responders* are, as expected, slightly more prone to recency effects than *slow responders*. This difference is, however, based only on one statistically significant recency effect for fast responders (see section A, dimension level "one sibling," $p = .036$).

4.5 Discussion and Conclusion

Our study adds a comprehensive experimental test of effects of dimension order in factorial survey experiments to the methodological literature on survey research. We asked whether the position of vignette dimensions in running text systematically influ-

Figure 4.2. Coefficient Plot of Recency Effects for the Full Sample and Subgroups

Note: This figure shows regression coefficients of two-way interaction terms of the relevant dimension level ([A] siblings, [B] work experience, [C] career orientation), with an indicator that the dimension was in the last position. Standard errors with 95% confidence intervals are attached. See Tables 4.D2 - 4.D8 in Appendix D for the regression tables.

ences the respondents' evaluations. We tested for *overall order effects* and for *recency effects* (i.e., whether, because of the complex structure of all the information to be retained and because of their proximity to the eventual evaluation task, respondents focus specifically on dimensions at the *end* of the running text). Moreover, in contrast to prior research that was based on student samples only, by using a general population sample, we were able to identify heterogeneous order effects across subgroups.

Following standard procedures, age, education, and response speed (as grouped variables) served as indicators of respondents' cognitive abilities. Overall, for our vignette module with four evaluations per respondent and seven different vignette dimensions, using a standard 11-point rating response scale, we found almost no systematic overall order effects. At best, there was weak evidence of recency effects within the group of fast responders. By age and education, only few and inconclusive effects can be reported, again with a slight tendency toward recency effects. In sum, the study supports the overall validity of factorial survey experiments with a medium degree of complexity

(seven dimensions are a typical design in this research; Wallander, 2009) in regard to the order in which vignette dimensions are presented.

Evidently, the study also has its limitations. First, the setup did not allow us to test for *primacy* effects because the first dimension was actually constant across all cases: All vignettes started with names indicating gender and ethnic background, for logical reasons. We indeed started variation at the second dimension, which is less salient in running texts than the first dimension. Second, we tested across seven dimensions, which is a typical number of dimensions used in applications. There are, however, applications using a higher number of dimensions, and there are good arguments that order effects most probably emerge only with more complex vignettes (Auspurg and Jäckle, 2017). Third, even more importantly, for obvious practical reasons, we tested only a subsample of possible vignette dimension orders. In all four treatment conditions, the attributes of the hypothetical applicants in the final position of the vignette texts only marginally influenced the substantive judgments in the evaluation task. One could argue that vignette dimensions with a higher substantive relevance, such as parental income and achievements at high school, as compared to number of siblings, career orientation, and work experience, could change the conclusions. This should be tested in future research.

Nevertheless, the study has its specific strengths in its rigorous tests based on data from a heterogeneous respondent population. The good news for practitioners of factorial survey experiments is that we did not find strong order effects of vignette dimensions at all. However, we also learned from the tests that vignette evaluations are obviously prone to some (small) recency effects. In designing factorial survey experiments, and particularly dimension order, researchers are well advised to consider a logical, realistic structure for their scenarios—with a specific focus on the dimension placed last. We suggest avoiding having dimensions with the highest supposed influence at the very end of running text.

The most challenging result of our study seems to be the sensitiveness of order effects to the differing speeds of responders. Those who take questionnaires quickly (who are probably "satisficers") are slightly more prone to possible bias caused by order effects. The use of online access panels is increasing, and respondents to these ubiquitous surveys are probably more likely to provide quick and imprecise responses, which might also invalidate results in other ways. Thus, another conclusion from our study points to the relevance of para-data when applying factorial survey experiments—and survey research in general. Such data might promote the elimination of possible bias caused by order effects and other effects related to satisficing behavior.

Appendices

Appendix A: Information on the Vignettes

Anja was born in Konstanz and has a high school diploma (GPA) of 1.3 ("very good").
Her parents have a mid-range income and no other children.
She's already completed a voluntary social year.

In her interview with the selection committee, she stated that she wanted to combine family and professional life as well as she could later on.

In your opinion, should this person rightfully receive a scholarship or not?

not at all										by all means
-5	-4	-3	-2	-1	0	1	2	3	4	5
□	□	□	□	□	□	□	□	□	□	□

Figure 4.A1. Sample Vignette for Treatment Group 1, With Underlined Dimensions

Note: The original German vignette was translated by the authors.

Table 4.A1. Vignette Dimensions and Levels

	Levels			
Dimensions	1	2	3	In total
Ethnical background	German	Foreign		2
Gender	Male	Female		2
Place of birth	Konstanz	Same province	Other German city	3
GPA	Very good (1.3)	Good (2.3)	Satisfactory (3.3)	3
Parents' income	Low	Middle	High	3
Siblings	None	One	Two	3
Work experience	Job training	Voluntary social year	no information provided	3
Career orientation	Career	Family & career	Family	3
			Vignette universe (2^2x3^6):	2,916

Note: GPA= grade point average

Table 4.A2. Information on Sample Size

	Four Order Groups				
	1	2	3	4	In total
Full sample:					
Number of respondents	188	215	180	204	787
Number of evaluations	749	852	709	809	3,119
Evaluations per subgroup:					
Under 60 yrs	598	668	561	631	2,478
60 yrs and older	151	164	148	178	641
Higher education	559	627	530	630	2,346
Lower education	190	225	179	179	773
Fast responders	358	449	354	404	1,565
Slow responders	391	403	355	405	1,554

Appendix B: Is there an Overall Order Effect?

Table 4.B1. Full Sample: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.255	(0.163)	−0.180	(0.171)	0.147	(0.167)	−0.073	(0.156)
Male (ref: female)	0.174	(0.171)	0.199	(0.176)	−0.135	(0.159)	0.026	(0.176)
Place of birth (ref: Konstanz)								
same province	0.091	(0.230)	0.062	(0.213)	0.447*	(0.218)	0.179	(0.202)
other German city	−0.053	(0.229)	0.145	(0.224)	0.070	(0.211)	0.236	(0.189)
GPA (ref: very good)								
good	−1.012***	(0.246)	−0.626***	(0.220)	−0.717***	(0.204)	−0.678***	(0.209)
satisfactory	−1.767***	(0.234)	−1.526***	(0.236)	−1.442***	(0.234)	−1.528***	(0.222)
Income parents (ref: low)								
middle	−1.439***	(0.209)	−1.192***	(0.212)	−1.272***	(0.228)	−1.165***	(0.193)
high	−4.157***	(0.243)	−4.024***	(0.243)	−4.146***	(0.233)	−4.469***	(0.247)
Siblings (ref: none)								
one	0.814***	(0.222)	0.334	(0.229)	1.113***	(0.187)	0.411*	(0.201)
two	0.908***	(0.230)	0.927***	(0.209)	1.084***	(0.216)	0.955***	(0.196)
Prev. work exp. (ref: job training)								
voluntary social year	0.721***	(0.229)	1.172***	(0.222)	0.535*	(0.225)	0.849***	(0.191)
no information provided	0.056	(0.227)	0.308	(0.214)	0.076	(0.231)	0.553**	(0.195)
Career orient. (ref: career)								
family & career	−0.121	(0.223)	−0.058	(0.224)	0.101	(0.206)	−0.401	(0.216)
family	−0.871***	(0.231)	−0.788***	(0.222)	−0.268	(0.211)	−0.622**	(0.225)
Constant	8.304***	(0.350)	7.775***	(0.379)	7.823***	(0.341)	8.213***	(0.316)
Observations	749		852		709		809	
R^2	0.352		0.324		0.367		0.389	

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.B2. 60 Years and Older: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.176	(0.329)	−0.020	(0.289)	−0.191	(0.455)	0.152	(0.337)
Male (ref: female)	0.653	(0.373)	1.095***	(0.346)	−0.174	(0.392)	−0.030	(0.452)
Place of birth (ref: Konstanz)								
same province	0.648	(0.595)	−0.455	(0.377)	0.525	(0.410)	0.114	(0.513)
other German city	0.005	(0.511)	0.554	(0.592)	0.730	(0.556)	0.525	(0.403)
GPA (ref: very good)								
good	−0.764	(0.574)	−0.654	(0.420)	−1.099*	(0.439)	−0.648	(0.520)
satisfactory	−1.328*	(0.558)	−0.585	(0.448)	−1.895***	(0.490)	−1.259*	(0.561)
Income parents (ref: low)								
middle	−0.976*	(0.466)	−2.084***	(0.446)	−1.801***	(0.575)	−1.819***	(0.448)
high	−3.635***	(0.508)	−4.618***	(0.476)	−4.349***	(0.563)	−5.394***	(0.607)
Siblings (ref: none)								
one	0.786	(0.521)	0.841*	(0.393)	0.732	(0.474)	0.822*	(0.396)
two	1.558***	(0.509)	0.552	(0.507)	1.664***	(0.467)	0.761	(0.497)
Prev. work exp. (ref: job training)								
voluntary social year	0.491	(0.641)	2.011***	(0.488)	0.704	(0.474)	0.585	(0.482)
no information provided	−0.458	(0.590)	0.574	(0.488)	0.373	(0.487)	0.716	(0.435)
Career orient. (ref: career)								
family & career	−0.261	(0.507)	−0.024	(0.512)	0.161	(0.487)	−0.706	(0.527)
family	−0.968	(0.597)	−0.493	(0.509)	−1.032*	(0.403)	−0.785	(0.547)
Constant	6.834***	(0.905)	7.441***	(0.801)	8.317***	(0.883)	8.370***	(1.044)
Observations	151		164		148		178	
R^2	0.327		0.440		0.452		0.449	

Note: Controlled for education. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.B3. Under 60 Years: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.260	(0.184)	−0.238	(0.198)	0.209	(0.180)	−0.158	(0.179)
Male (ref: female)	0.077	(0.197)	0.009	(0.197)	−0.178	(0.174)	−0.011	(0.194)
Place of birth (ref: Konstanz)								
same province	0.040	(0.246)	0.215	(0.243)	0.456	(0.249)	0.228	(0.218)
other German city	−0.022	(0.262)	0.139	(0.236)	−0.090	(0.225)	0.160	(0.217)
GPA (ref: very good)								
good	−1.075***	(0.279)	−0.654*	(0.256)	−0.665***	(0.231)	−0.638***	(0.215)
satisfactory	−1.829***	(0.266)	−1.699***	(0.264)	−1.328***	(0.263)	−1.583***	(0.244)
Income parents (ref: low)								
middle	−1.517***	(0.239)	−1.046***	(0.243)	−1.167***	(0.258)	−1.000***	(0.214)
high	−4.211***	(0.282)	−3.924***	(0.275)	−4.168***	(0.267)	−4.242***	(0.265)
Siblings (ref: none)								
one	0.835***	(0.260)	0.271	(0.263)	1.151***	(0.210)	0.312	(0.232)
two	0.768***	(0.259)	1.018***	(0.232)	0.924***	(0.247)	1.000***	(0.216)
Prev. work exp. (ref: job training)								
voluntary social year	0.856***	(0.248)	1.004***	(0.247)	0.438	(0.269)	0.986***	(0.208)
no information provided	0.157	(0.253)	0.216	(0.233)	−0.059	(0.262)	0.452*	(0.221)
Career orient. (ref: career)								
family & career	−0.042	(0.255)	0.044	(0.252)	0.111	(0.234)	−0.262	(0.250)
family	−0.804***	(0.264)	−0.836***	(0.260)	−0.070	(0.244)	−0.523*	(0.252)
Constant	8.133***	(0.501)	7.497***	(0.507)	8.080***	(0.460)	7.956***	(0.433)
Observations	598		688		561		631	
R^2	0.371		0.322		0.367		0.387	

Note: Controlled for education. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.B4. Higher Education Level: OLS-Regression of Vignette Evaluation on Vignette Dimension by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.325	(0.189)	−0.229	(0.205)	0.150	(0.186)	0.053	(0.174)
Male (ref: female)	0.045	(0.200)	0.074	(0.210)	−0.130	(0.183)	0.061	(0.194)
Place of birth (ref: Konstanz)								
same province	0.184	(0.258)	0.381	(0.244)	0.513*	(0.241)	0.269	(0.232)
other German city	−0.156	(0.271)	0.302	(0.269)	0.084	(0.253)	0.096	(0.206)
GPA (ref: very good)								
good	−1.116***	(0.292)	−0.579*	(0.254)	−0.729***	(0.237)	−0.726***	(0.223)
satisfactory	−1.676***	(0.265)	−1.449***	(0.290)	−1.493***	(0.280)	−1.673***	(0.251)
Income parents (ref: low)								
middle	−1.394***	(0.244)	−1.241***	(0.243)	−1.504***	(0.256)	−1.212***	(0.213)
high	−4.288***	(0.283)	−3.998***	(0.284)	−4.224***	(0.259)	−4.420***	(0.278)
Siblings (ref: none)								
one	0.716**	(0.255)	0.296	(0.278)	1.103***	(0.218)	0.462*	(0.226)
two	0.861***	(0.249)	0.995***	(0.243)	1.075***	(0.256)	0.992***	(0.216)
Prev. work exp. (ref: job training)								
voluntary social year	0.988***	(0.266)	1.400***	(0.263)	0.405	(0.267)	1.012***	(0.201)
no information provided	0.194	(0.256)	0.229	(0.259)	0.068	(0.274)	0.704***	(0.222)
Career orient. (ref: career)								
family & career	−0.253	(0.259)	−0.056	(0.260)	−0.003	(0.230)	−0.315	(0.234)
family	−1.072***	(0.263)	−0.826***	(0.268)	−0.279	(0.252)	−0.592*	(0.257)
Constant	8.481***	(0.427)	7.711***	(0.448)	7.811***	(0.410)	7.937***	(0.343)
Observations	559		627		530		630	
R^2	0.387		0.331		0.383		0.410	

Note: Controlled for age. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.B5. Lower Education Level: OLS-Regression of Vignette Evaluation on Vignette Dimension by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.081	(0.297)	−0.149	(0.325)	−0.053	(0.394)	−0.486	(0.374)
Male (ref: female)	0.671*	(0.332)	0.652	(0.351)	−0.107	(0.311)	−0.094	(0.405)
Place of birth (ref: Konstanz)								
same province	−0.285	(0.507)	−0.883	(0.450)	0.394	(0.489)	−0.182	(0.456)
other German city	0.207	(0.461)	−0.255	(0.446)	0.100	(0.423)	0.709	(0.439)
GPA (ref: very good)								
good	−0.973*	(0.442)	−0.655	(0.414)	−0.528	(0.471)	−0.615	(0.550)
satisfactory	−2.305***	(0.455)	−1.720***	(0.442)	−1.268**	(0.447)	−1.044	(0.567)
Income parents (ref: low)								
middle	−1.705***	(0.446)	−1.279**	(0.450)	−0.684	(0.606)	−1.097**	(0.399)
high	−3.869***	(0.467)	−4.258***	(0.482)	−3.981***	(0.517)	−4.521***	(0.561)
Siblings (ref: none)								
one	1.164**	(0.426)	0.488	(0.438)	1.214**	(0.420)	0.250	(0.427)
two	0.911	(0.548)	0.743	(0.412)	0.947*	(0.427)	1.099*	(0.497)
Prev. work exp. (ref: job training)								
voluntary social year	−0.027	(0.454)	0.736	(0.408)	0.756	(0.458)	0.510	(0.554)
no information provided	−0.434	(0.478)	0.630	(0.403)	0.091	(0.432)	0.031	(0.463)
Career orient. (ref: career)								
family & career	0.566	(0.447)	−0.027	(0.456)	0.450	(0.499)	−0.717	(0.503)
family	−0.243	(0.495)	−0.739	(0.435)	−0.076	(0.375)	−1.006	(0.512)
Constant	8.002***	(0.635)	7.801***	(0.835)	7.557***	(0.723)	8.664***	(0.816)
Observations	190		225		179		179	
R^2	0.310		0.359		0.360		0.382	

Note: Controlled for age. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.B6. Fast Responders: OLS-Regression of Vignette Evaluation on Vignette Dimension by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.526	(0.270)	−0.164	(0.218)	−0.031	(0.200)	−0.029	(0.212)
Male (ref: female)	−0.146	(0.253)	0.501	(0.258)	0.150	(0.206)	−0.017	(0.256)
Place of birth (ref: Konstanz)								
same province	−0.106	(0.328)	−0.180	(0.301)	0.166	(0.290)	−0.035	(0.254)
other German city	−0.023	(0.365)	−0.264	(0.292)	0.004	(0.261)	0.302	(0.273)
GPA (ref: very good)								
good	−0.976*	(0.379)	−0.619	(0.317)	−0.779**	(0.280)	−0.284	(0.310)
satisfactory	−1.696***	(0.357)	−1.586***	(0.355)	−1.314***	(0.307)	−1.098***	(0.316)
Income parents (ref: low)								
middle	−1.585***	(0.289)	−0.991***	(0.273)	−0.980***	(0.311)	−1.011***	(0.260)
high	−4.095***	(0.382)	−4.220***	(0.334)	−4.044***	(0.344)	−4.404***	(0.332)
Siblings (ref: none)								
one	1.230***	(0.324)	0.486	(0.292)	1.327***	(0.235)	0.395	(0.286)
two	0.971**	(0.350)	1.008***	(0.297)	0.760**	(0.274)	0.894***	(0.308)
Prev. work exp. (ref: job training)								
voluntary social year	1.138***	(0.356)	1.374***	(0.302)	0.394	(0.292)	0.735**	(0.275)
no information provided	0.102	(0.312)	0.277	(0.303)	0.032	(0.309)	0.409	(0.278)
Career orient. (ref: career)								
family & career	0.031	(0.326)	0.023	(0.317)	−0.090	(0.239)	−0.452	(0.302)
family	−0.976**	(0.344)	−0.989***	(0.323)	−0.241	(0.277)	−0.612	(0.322)
Constant	8.093***	(0.567)	7.857***	(0.527)	7.997***	(0.441)	8.109***	(0.457)
Observations	358		449		354		404	
R^2	0.365		0.349		0.375		0.380	

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.B7. Slow Responders: OLS-Regression of Vignette Evaluation on Vignette Dimension by Treatment Split

	Order 1		Order 2		Order 3		Order 4	
Foreign (ref: German)	0.029	(0.210)	−0.172	(0.268)	0.402	(0.277)	−0.050	(0.230)
Male (ref: female)	0.462	(0.239)	−0.037	(0.243)	−0.367	(0.232)	0.103	(0.243)
Place of birth (ref: Konstanz)								
same province	0.284	(0.320)	0.273	(0.300)	0.756*	(0.318)	0.332	(0.310)
other German city	−0.082	(0.299)	0.484	(0.348)	0.174	(0.327)	0.128	(0.262)
GPA (ref: very good)								
good	−1.061***	(0.334)	−0.715*	(0.319)	−0.617*	(0.295)	−1.056***	(0.282)
satisfactory	−1.766***	(0.338)	−1.538***	(0.317)	−1.558***	(0.381)	−1.932***	(0.310)
Income parents (ref: low)								
middle	−1.270***	(0.292)	−1.368***	(0.324)	−1.529***	(0.322)	−1.314***	(0.294)
high	−4.187***	(0.314)	−3.795***	(0.373)	−4.177***	(0.316)	−4.575***	(0.370)
Siblings (ref: none)								
one	0.373	(0.290)	0.274	(0.355)	0.912***	(0.304)	0.386	(0.283)
two	0.862**	(0.305)	0.802**	(0.294)	1.439***	(0.334)	1.027***	(0.251)
Prev. work exp. (ref: job training)								
voluntary social year	0.303	(0.288)	1.022***	(0.323)	0.615	(0.324)	0.950***	(0.269)
no information provided	−0.003	(0.327)	0.369	(0.315)	0.119	(0.333)	0.721*	(0.276)
Career orient. (ref: career)								
family & career	−0.178	(0.301)	−0.140	(0.331)	0.345	(0.332)	−0.341	(0.314)
family	−0.691*	(0.302)	−0.594	(0.309)	−0.223	(0.331)	−0.667*	(0.328)
Constant	8.434***	(0.456)	7.663***	(0.558)	7.471***	(0.533)	8.321***	(0.452)
Observations	391		403		355		405	
R^2	0.364		0.311		0.383		0.410	

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Appendix C: Individual Wald Chi^2 Tests for the Full Sample and all Subgroups

Table 4.C1. Individual Wald Chi^2 Tests for the Full Sample

Comparison groups	Dimension	Chi^2	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	0.265	.606	n.s.	
	Other German city	0.001	.997	n.s.	
	GPA (ref: very good)				
	Good	0.045	.832	n.s.	
	Satisfactory	0.042	.837	n.s.	
	Parent's income (ref: low)				
	Middle	1.006	.316	n.s.	
	High	0.400	.527	n.s.	
	Siblings (ref: none)				
	One	2.824	.093	+	no order eff.
	Two	0.040	.842	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.813	.178	n.s.	
	Career orientation (ref: career)				
	Family & career	0.137	.711	n.s	
	Family	0.006	.936	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.259	.611	n.s.	
	Other German city	0.045	.831	n.s.	
	GPA (ref: very good)				
	Good	0.533	.465	n.s.	
	Satisfactory	0.204	.651	n.s.	
	Parent's income (ref: low)				
	Middle	0.788	.375	n.s.	
	High	0.137	.712	n.s.	
	Siblings (ref: none)				
	One	0.643	.442	n.s.	
	Two	0.317	.573	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.479	.489	n.s.	
	Career orientation (ref: career)				
	Family & career	0.000	.987	n.s	
	Family	0.833	.361	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.630	.427	n.s.	
	Other German city	0.393	.531	n.s.	
	GPA (ref: very good)				
	Good	0.068	.794	n.s.	
	Satisfactory	0.580	.446	n.s.	
	Parent's income (ref: low)				
	Middle	0.008	.930	n.s.	
	High	0.125	.724	n.s.	
	Siblings (ref: none)				
	One	1.270	.260	n.s.	
	Two	0.161	.688	n.s.	

Continued on next page.

Table 4.C1 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Work experience (ref: job training)				
	Vol. social year	0.243	.662	n.s.	
	Career orientation (ref: career)				
	Family & career	0.438	.508	n.s	
	Family	0.570	.450	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.000	.997	n.s.	
	Other German city	0.062	.804	n.s.	
	GPA (ref: very good)				
	Good	0.316	.574	n.s.	
	Satisfactory	0.427	.513	n.s.	
	Parent's income (ref: low)				
	Middle	0.014	.905	n.s.	
	High	0.083	.773	n.s.	
	Siblings (ref: none)				
	One	6.987	.008	**	Recency
	Two	0.162	.687	n.s.	
	Work experience (ref: job training)				
	Vol. social year	4.175	.041	*	Recency
	Career orientation (ref: career)				
	Family & career	0.142	.707	n.s	
	Family	1.048	.306	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.086	.770	n.s.	
	Other German city	0.376	.540	n.s.	
	GPA (ref: very good)				
	Good	0.003	.957	n.s.	
	Satisfactory	0.293	.588	n.s.	
	Parent's income (ref: low)				
	Middle	1.362	.343	n.s.	
	High	0.959	.328	n.s.	
	Siblings (ref: none)				
	One	0.394	.530	n.s.	
	Two	0.048	.826	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.908	.341	n.s.	
	Career orientation (ref: career)				
	Family & career	0.091	.763	n.s	
	Family	0.743	.384	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.087	.768	n.s.	
	Other German city	0.777	.378	n.s.	
	GPA (ref: very good)				
	Good	0.258	.611	n.s.	
	Satisfactory	1.526	.217	n.s.	
	Parent's income (ref: low)				
	Middle	1.089	.297	n.s.	
	High	0.538	.463	n.s.	
	Siblings (ref: none)				
	One	4.349	.037	*	Recency

Continued on next page.

Table 4.C1 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Two	0.039	.844	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.530	.216	n.s.	
	Career orientation (ref: career)				
	Family & career	0.472	.492	n.s	
	Family	0.029	.865	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi^2 tests. Each Chi^2 stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.C2. Individual Wald Chi² Tests for Subgroup: 60 Years and Older

Comparison groups	Dimension	Chi²	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	0.805	.370	n.s.	
	Other German city	0.489	.484	n.s.	
	GPA (ref: very good)				
	Good	0.094	.759	n.s.	
	Satisfactory	0.195	.659	n.s.	
	Parent's income (ref: low)				
	Middle	1.743	.187	n.s.	
	High	1.466	.226	n.s.	
	Siblings (ref: none)				
	One	0.064	.801	n.s.	
	Two	1.250	.264	n.s.	
	Work experience (ref: job training)				
	Vol. social year	4.658	.031	*	Recency
	Career orientation (ref: career)				
	Family & career	0.512	.474	n.s	
	Family	0.338	.561	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.037	.847	n.s.	
	Other German city	1.595	.207	n.s.	
	GPA (ref: very good)				
	Good	0.003	.960	n.s.	
	Satisfactory	0.491	.483	n.s.	
	Parent's income (ref: low)				
	Middle	0.776	.378	n.s.	
	High	0.614	.433	n.s.	
	Siblings (ref: none)				
	One	0.120	.729	n.s.	
	Two	0.183	.669	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.251	.616	n.s.	
	Career orientation (ref: career)				
	Family & career	0.193	.660	n.s	
	Family	0.139	.710	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.070	.791	n.s.	
	Other German city	0.066	.797	n.s.	
	GPA (ref: very good)				
	Good	0.050	.822	n.s.	
	Satisfactory	0.054	.817	n.s.	
	Parent's income (ref: low)				
	Middle	0.398	.528	n.s.	
	High	3.503	.061	+	Recency
	Siblings (ref: none)				
	One	0.001	.974	n.s.	
	Two	0.247	.619	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.057	.812	n.s.	
	Career orientation (ref: career)				

Continued on next page.

Table 4.C2 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Family & career	0.075	.784	n.s	
	Family	0.005	.942	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	2.349	.125	n.s.	
	Other German city	0.299	.584	n.s.	
	GPA (ref: very good)				
	Good	0.088	.767	n.s.	
	Satisfactory	1.840	.175	n.s.	
	Parent's income (ref: low)				
	Middle	0.177	.674	n.s.	
	High	0.224	.636	n.s.	
	Siblings (ref: none)				
	One	0.015	.904	n.s.	
	Two	0.706	.401	n.s.	
	Work experience (ref: job training)				
	Vol. social year	4.169	.041	*	Recency
	Career orientation (ref: career)				
	Family & career	0.091	.763	n.s	
	Family	1.276	.259	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.378	.539	n.s.	
	Other German city	1.091	.296	n.s.	
	GPA (ref: very good)				
	Good	0.379	.538	n.s.	
	Satisfactory	0.024	.876	n.s.	
	Parent's income (ref: low)				
	Middle	0.327	.568	n.s.	
	High	0.455	.500	n.s.	
	Siblings (ref: none)				
	One	0.067	.795	n.s.	
	Two	0.500	.479	n.s.	
	Work experience (ref: job training)				
	Vol. social year	4.715	.030	*	Recency
	Career orientation (ref: career)				
	Family & career	1.033	.309	n.s	
	Family	0.572	.450	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.280	.596	n.s.	
	Other German city	2.872	.090	+	Recency
	GPA (ref: very good)				
	Good	0.094	.760	n.s.	
	Satisfactory	0.836	.360	n.s.	
	Parent's income (ref: low)				
	Middle	0.034	.853	n.s.	
	High	1.354	.245	n.s.	
	Siblings (ref: none)				
	One	0.135	.713	n.s.	
	Two	0.010	.921	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.082	.775	n.s.	

Continued on next page.

Table 4.C2 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Career orientation (ref: career)				
	Family & career	0.549	.459	n.s	
	Family	0.111	.739	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi² tests. Each Chi² stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.C3. Individual Wald Chi^2 Tests for Subgroup: Under 60 Years

Comparison groups	Dimension	Chi^2	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	0.977	.323	n.s.	
	Other German city	0.002	.966	n.s.	
	GPA (ref: very good)				
	Good	0.011	.916	n.s.	
	Satisfactory	0.122	.727	n.s.	
	Parent's income (ref: low)				
	Middle	2.208	.137	n.s.	
	High	1.270	.260	n.s.	
	Siblings (ref: none)				
	One	3.076	.079	+	no order eff.
	Two	0.424	.515	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.059	.808	n.s.	
	Career orientation (ref: career)				
	Family & career	0.262	.609	n.s	
	Family	0.123	.726	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.250	.617	n.s.	
	Other German city	0.699	.403	n.s.	
	GPA (ref: very good)				
	Good	0.819	.366	n.s.	
	Satisfactory	0.710	.400	n.s.	
	Parent's income (ref: low)				
	Middle	2.177	.140	n.s.	
	High	0.456	.499	n.s.	
	Siblings (ref: none)				
	One	0.439	.507	n.s.	
	Two	0.572	.449	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.807	.179	n.s.	
	Career orientation (ref: career)				
	Family & career	0.039	.843	n.s	
	Family	1.552	.213	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.858	.354	n.s.	
	Other German city	0.431	.512	n.s.	
	GPA (ref: very good)				
	Good	0.186	.666	n.s.	
	Satisfactory	1.625	.202	n.s.	
	Parent's income (ref: low)				
	Middle	0.044	.834	n.s.	
	High	0.241	.623	n.s.	
	Siblings (ref: none)				
	One	1.657	.198	n.s.	
	Two	0.642	.423	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.099	.753	n.s.	
	Career orientation (ref: career)				

Continued on next page.

Table 4.C3 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Family & career	0.393	.531	n.s	
	Family	0.637	.425	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.243	.622	n.s.	
	Other German city	0.694	.405	n.s.	
	GPA (ref: very good)				
	Good	0.733	.392	n.s.	
	Satisfactory	1.336	.248	n.s.	
	Parent's income (ref: low)				
	Middle	0.000	.996	n.s.	
	High	0.229	.632	n.s.	
	Siblings (ref: none)				
	One	6.575	.010	*	Recency
	Two	0.022	.881	n.s.	
	Work experience (ref: job training)				
	Vol. social year	2.461	.117	n.s.	
	Career orientation (ref: career)				
	Family & career	0.109	.742	n.s	
	Family	2.634	.105	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.009	.926	n.s.	
	Other German city	0.551	.458	n.s.	
	GPA (ref: very good)				
	Good	0.120	.729	n.s.	
	Satisfactory	0.743	.389	n.s.	
	Parent's income (ref: low)				
	Middle	2.004	.157	n.s.	
	High	0.487	.458	n.s.	
	Siblings (ref: none)				
	One	0.314	.575	n.s.	
	Two	0.022	.883	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.003	.959	n.s.	
	Career orientation (ref: career)				
	Family & career	0.013	.910	n.s	
	Family	1.369	.242	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.173	.678	n.s.	
	Other German city	2.668	.102	n.s.	
	GPA (ref: very good)				
	Good	0.288	.592	n.s.	
	Satisfactory	4.580	.032	*	Recency
	Parent's income (ref: low)				
	Middle	1.970	.160	n.s.	
	High	0.044	.834	n.s.	
	Siblings (ref: none)				
	One	4.460	.035	*	Recency
	Two	0.000	.988	n.s.	
	Work experience (ref: job training)				
	Vol. social year	2.995	.084	+	no order eff.

Continued on next page.

Table 4.C3 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Career orientation (ref: career)				
	Family & career	0.201	.654	n.s	
	Family	0.216	.642	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi^2 tests. Each Chi^2 stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.C4. Individual Wald Chi² Tests for Subgroup: Higher Education Level

Comparison groups	Dimension	Chi²	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	1.870	.172	n.s.	
	Other German city	0.789	.374	n.s.	
	GPA (ref: very good)				
	Good	0.930	.335	n.s.	
	Satisfactory	0.162	.687	n.s.	
	Parent's income (ref: low)				
	Middle	0.251	.616	n.s.	
	High	1.304	.254	n.s.	
	Siblings (ref: none)				
	One	1.693	.193	n.s.	
	Two	0.415	.519	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.962	.327	n.s.	
	Career orientation (ref: career)				
	Family & career	0.151	.697	n.s	
	Family	0.343	.558	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.536	.464	n.s.	
	Other German city	0.010	.921	n.s.	
	GPA (ref: very good)				
	Good	1.203	.273	n.s.	
	Satisfactory	0.002	.969	n.s.	
	Parent's income (ref: low)				
	Middle	0.067	.796	n.s.	
	High	0.204	.651	n.s.	
	Siblings (ref: none)				
	One	0.920	.338	n.s.	
	Two	0.278	.598	n.s.	
	Work experience (ref: job training)				
	Vol. social year	2.686	.101	n.s.	
	Career orientation (ref: career)				
	Family & career	0.194	.660	n.s	
	Family	1.699	.192	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	1.418	.234	n.s.	
	Other German city	0.126	.722	n.s.	
	GPA (ref: very good)				
	Good	0.374	.541	n.s.	
	Satisfactory	1.861	.172	n.s.	
	Parent's income (ref: low)				
	Middle	0.437	.509	n.s.	
	High	0.018	.892	n.s.	
	Siblings (ref: none)				
	One	0.374	.541	n.s.	
	Two	0.421	.516	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.005	.941	n.s.	
	Career orientation (ref: career)				

Continued on next page.

Table 4.C4 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Family & career	0.061	.804	n.s	
	Family	1.15	.284	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.446	.504	n.s.	
	Other German city	0.631	.427	n.s.	
	GPA (ref: very good)				
	Good	0.024	.878	n.s.	
	Satisfactory	1.134	.714	n.s.	
	Parent's income (ref: low)				
	Middle	0.064	.800	n.s.	
	High	0.524	.469	n.s.	
	Siblings (ref: none)				
	One	5.626	.018	*	Recency
	Two	0.008	.929	n.s.	
	Work experience (ref: job training)				
	Vol. social year	6.741	.009	**	Recency
	Career orientation (ref: career)				
	Family & career	0.003	.955	n.s	
	Family	0.478	.489	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.046	.830	n.s.	
	Other German city	0.441	.507	n.s.	
	GPA (ref: very good)				
	Good	0.175	.676	n.s.	
	Satisfactory	0.780	.377	n.s.	
	Parent's income (ref: low)				
	Middle	1.600	.206	n.s.	
	High	1.624	.203	n.s.	
	Siblings (ref: none)				
	One	0.592	.442	n.s.	
	Two	0.000	.984	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.034	.309	n.s.	
	Career orientation (ref: career)				
	Family & career	0.028	.866	n.s	
	Family	0.202	.653	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.219	.640	n.s.	
	Other German city	0.06	.807	n.s.	
	GPA (ref: very good)				
	Good	0.328	.567	n.s.	
	Satisfactory	1.767	.184	n.s.	
	Parent's income (ref: low)				
	Middle	0.979	.332	n.s.	
	High	0.349	.555	n.s.	
	Siblings (ref: none)				
	One	2.875	.090	+	Recency
	Two	0.005	.942	n.s.	
	Work experience (ref: job training)				
	Vol. social year	3.596	.058	+	no order eff.

Continued on next page.

Table 4.C4 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Career orientation (ref: career)				
	Family & career	0.051	.821	n.s	
	Family	0.071	.791	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi² tests. Each Chi² stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.C5. Individual Wald Chi^2 Tests for Subgroup: Lower Education Level

Comparison groups	Dimension	Chi^2	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	0.478	.490	n.s.	
	Other German city	0.961	.327	n.s.	
	GPA (ref: very good)				
	Good	0.147	.701	n.s.	
	Satisfactory	0.700	.403	n.s.	
	Parent's income (ref: low)				
	Middle	0.865	.352	n.s.	
	High	0.516	.472	n.s.	
	Siblings (ref: none)				
	One	1.226	.268	n.s.	
	Two	0.042	.837	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.521	.218	n.s.	
	Career orientation (ref: career)				
	Family & career	1.970	.160	n.s	
	Family	1.025	.311	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.172	.678	n.s.	
	Other German city	0.029	.865	n.s.	
	GPA (ref: very good)				
	Good	0.143	.705	n.s.	
	Satisfactory	2.406	.121	n.s.	
	Parent's income (ref: low)				
	Middle	1.675	.196	n.s.	
	High	0.011	.916	n.s.	
	Siblings (ref: none)				
	One	0.001	.981	n.s.	
	Two	0.120	.729	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.169	.280	n.s.	
	Career orientation (ref: career)				
	Family & career	1.171	.279	n.s	
	Family	0.037	.847	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.002	.962	n.s.	
	Other German city	0.333	.564	n.s.	
	GPA (ref: very good)				
	Good	0.005	.945	n.s.	
	Satisfactory	1.284	.257	n.s.	
	Parent's income (ref: low)				
	Middle	0.205	.650	n.s.	
	High	0.403	.525	n.s.	
	Siblings (ref: none)				
	One	1.464	.226	n.s.	
	Two	0.647	.421	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.222	.638	n.s.	
	Career orientation (ref: career)				

Continued on next page.

Table 4.C5 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Family & career	4.421	.036	*	no order eff.
	Family	0.709	.400	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	1.343	.247	n.s.	
	Other German city	0.798	.370	n.s.	
	GPA (ref: very good)				
	Good	0.549	.459	n.s.	
	Satisfactory	0.506	.477	n.s.	
	Parent's income (ref: low)				
	Middle	0.184	.668	n.s.	
	High	0.418	.518	n.s.	
	Siblings (ref: none)				
	One	1.536	.215	n.s.	
	Two	0.430	.512	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.004	.950	n.s.	
	Career orientation (ref: career)				
	Family & career	0.171	.679	n.s	
	Family	0.950	.330	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.384	.535	n.s.	
	Other German city	2.814	.093	+	Primacy
	GPA (ref: very good)				
	Good	0.171	.680	n.s.	
	Satisfactory	0.236	.627	n.s.	
	Parent's income (ref: low)				
	Middle	0.288	.592	n.s.	
	High	0.002	.962	n.s.	
	Siblings (ref: none)				
	One	0.017	.896	n.s.	
	Two	1.277	.259	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.336	.562	n.s.	
	Career orientation (ref: career)				
	Family & career	0.553	.457	n.s	
	Family	0.008	.929	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.205	.651	n.s.	
	Other German city	0.677	.410	n.s.	
	GPA (ref: very good)				
	Good	0.075	.784	n.s.	
	Satisfactory	0.002	.965	n.s.	
	Parent's income (ref: low)				
	Middle	0.897	.334	n.s.	
	High	0.315	.575	n.s.	
	Siblings (ref: none)				
	One	1.813	.178	n.s.	
	Two	0.307	.580	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.244	.622	n.s.	

Continued on next page.

Table 4.C5 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Career orientation (ref: career)				
	Family & career	1.404	.236	n.s	
	Family	0.587	.444	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi^2 tests. Each Chi^2 stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.C6. Individual Wald Chi² Tests for Subgroup: Fast Responders

Comparison groups	Dimension	Chi²	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	0.053	.820	n.s.	
	Other German city	0.380	.538	n.s.	
	GPA (ref: very good)				
	Good	0.000	.999	n.s.	
	Satisfactory	1.098	.295	n.s.	
	Parent's income (ref: low)				
	Middle	3.124	.077	+	no order eff.
	High	0.016	.901	n.s.	
	Siblings (ref: none)				
	One	6.015	.014	*	no order eff.
	Two	0.024	.876	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.128	.720	n.s.	
	Career orientation (ref: career)				
	Family & career	0.024	.876	n.s	
	Family	0.001	.974	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.12	.729	n.s.	
	Other German city	0.021	.884	n.s.	
	GPA (ref: very good)				
	Good	0.525	.469	n.s.	
	Satisfactory	0.014	.907	n.s.	
	Parent's income (ref: low)				
	Middle	2.571	.109	n.s.	
	High	0.038	.846	n.s.	
	Siblings (ref: none)				
	One	0.003	.957	n.s.	
	Two	0.010	.919	n.s.	
	Work experience (ref: job training)				
	Vol. social year	3.422	.064	+	Recency
	Career orientation (ref: career)				
	Family & career	0.561	.454	n.s	
	Family	0.563	.453	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.353	.553	n.s.	
	Other German city	0.106	.745	n.s.	
	GPA (ref: very good)				
	Good	1.802	.179	n.s.	
	Satisfactory	0.066	.797	n.s.	
	Parent's income (ref: low)				
	Middle	0.574	.449	n.s.	
	High	0.537	.464	n.s.	
	Siblings (ref: none)				
	One	4.433	.035	*	Primacy
	Two	0.004	.947	n.s.	
	Work experience (ref: job training)				
	Vol. social year	1.133	.287	n.s.	
	Career orientation (ref: career)				

Continued on next page.

Table 4.C6 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Family & career	0.434	.510	n.s.	
	Family	0.495	.482	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.015	.904	n.s.	
	Other German city	0.348	.555	n.s.	
	GPA (ref: very good)				
	Good	0.722	.396	n.s.	
	Satisfactory	1.047	.306	n.s.	
	Parent's income (ref: low)				
	Middle	0.028	.867	n.s.	
	High	0.104	.747	n.s.	
	Siblings (ref: none)				
	One	8.018	.005	**	Recency
	Two	0.005	.946	n.s.	
	Work experience (ref: job training)				
	Vol. social year	5.861	.015	*	Recency
	Career orientation (ref: career)				
	Family & career	0.394	.530	n.s	
	Family	0.679	.410	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.157	.692	n.s.	
	Other German city	1.314	.252	n.s.	
	GPA (ref: very good)				
	Good	2.385	.123	n.s.	
	Satisfactory	1.817	.178	n.s.	
	Parent's income (ref: low)				
	Middle	1.274	.259	n.s.	
	High	0.34	.560	n.s.	
	Siblings (ref: none)				
	One	0.162	.687	n.s.	
	Two	0.010	.921	n.s.	
	Work experience (ref: job training)				
	Vol. social year	2.475	.116	n.s.	
	Career orientation (ref: career)				
	Family & career	0.283	.595	n.s	
	Family	0.595	.440	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.086	.770	n.s.	
	Other German city	0.361	.548	n.s.	
	GPA (ref: very good)				
	Good	0.626	.429	n.s.	
	Satisfactory	0.163	.687	n.s.	
	Parent's income (ref: low)				
	Middle	0.905	.341	n.s.	
	High	1.003	.317	n.s.	
	Siblings (ref: none)				
	One	5.917	.015	*	Recency
	Two	0.001	.971	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.833	.361	n.s.	

Continued on next page.

Table 4.C6 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Career orientation (ref: career)				
	Family & career	0.011	.917	n.s	
	Family	0.001	.979	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi² tests. Each Chi² stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 4.C7. Individual Wald Chi^2 Tests for Subgroup: Slow Responders

Comparison groups	Dimension	Chi^2	p	sig.	Type
order 1 vs. 2	*Place of birth (ref: Konstanz)*				
	Same province	0.441	.506	n.s.	
	Other German city	0.74	.390	n.s.	
	GPA (ref: very good)				
	Good	0.046	.831	n.s.	
	Satisfactory	0.274	.601	n.s.	
	Parent's income (ref: low)				
	Middle	0.037	.848	n.s	
	High	1.232	.267	n.s.	
	Siblings (ref: none)				
	One	0.324	.569	n.s	
	Two	0.411	.521	n.s.	
	Work experience (ref: job training)				
	Vol. social year	3.054	.081	+	Recency
	Career orientation (ref: career)				
	Family & career	0.155	.693	n.s	
	Family	0.000	.998	n.s.	
order 1 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.094	.759	n.s.	
	Other German city	0.000	1.000	n.s.	
	GPA (ref: very good)				
	Good	0.058	.809	n.s.	
	Satisfactory	0.249	.618	n.s.	
	Parent's income (ref: low)				
	Middle	0.060	.806	n.s.	
	High	0.101	.751	n.s.	
	Siblings (ref: none)				
	One	2.599	.107	n.s.	
	Two	1.152	.283	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.675	.411	n.s	
	Career orientation (ref: career)				
	Family & career	0.385	.535	n.s	
	Family	0.276	.599	n.s.	
order 1 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.114	.736	n.s.	
	Other German city	0.289	.591	n.s.	
	GPA (ref: very good)				
	Good	1.269	.260	n.s.	
	Satisfactory	2.564	.109	n.s.	
	Parent's income (ref: low)				
	Middle	0.714	.398	n.s.	
	High	0.015	.903	n.s.	
	Siblings (ref: none)				
	One	0.407	.523	n.s.	
	Two	0.587	.444	n.s.	
	Work experience (ref: job training)				
	Vol. social year	3.216	.073	+	Primacy
	Career orientation (ref: career)				

Continued on next page.

Table 4.C7 – *Continued from previous page.*

Comparison groups	Dimension	Chi²	p	sig.	Type
	Family & career	0.158	.691	n.s.	
	Family	0.013	.909	n.s.	
order 2 vs. 3	*Place of birth (ref: Konstanz)*				
	Same province	0.098	.754	n.s.	
	Other German city	0.567	.452	n.s.	
	GPA (ref: very good)				
	Good	0.001	.978	n.s.	
	Satisfactory	0.000	.995	n.s.	
	Parent's income (ref: low)				
	Middle	0.002	.960	n.s.	
	High	0.585	.444	n.s.	
	Siblings (ref: none)				
	One	0.959	.327	n.s.	
	Two	0.280	.597	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.739	.390	n.s.	
	Career orientation (ref: career)				
	Family & career	0.958	.328	n.s	
	Family	0.275	.600	n.s.	
order 2 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.115	.735	n.s.	
	Other German city	0.149	.699	n.s.	
	GPA (ref: very good)				
	Good	1.943	.163	n.s	
	Satisfactory	4.453	.035	*	Recency
	Parent's income (ref: low)				
	Middle	0.415	.519	n.s.	
	High	0.764	.382	n.s.	
	Siblings (ref: none)				
	One	0.002	.969	n.s.	
	Two	0.017	.896	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.008	.927	n.s.	
	Career orientation (ref: career)				
	Family & career	0.000	.994	n.s	
	Family	0.013	.908	n.s.	
order 3 vs. 4	*Place of birth (ref: Konstanz)*				
	Same province	0.000	.998	n.s.	
	Other German city	0.208	.648	n.s.	
	GPA (ref: very good)				
	Good	2.044	.153	n.s.	
	Satisfactory	4.124	.042	*	Recency
	Parent's income (ref: low)				
	Middle	0.369	.543	n.s.	
	High	0.027	.870	n.s.	
	Siblings (ref: none)				
	One	0.964	.326	n.s.	
	Two	0.173	.677	n.s.	
	Work experience (ref: job training)				
	Vol. social year	0.687	.407	n.s.	

Continued on next page.

Table 4.C7 – *Continued from previous page.*

Comparison groups	Dimension	Chi^2	p	sig.	Type
	Career orientation (ref: career)				
	Family & career	0.952	.329	n.s	
	Family	0.167	.683	n.s.	

Note: The table compares effect sizes of the same dimension between different treatment groups by using Chi^2 tests. Each Chi^2 stands for an individual Wald test that compares the relevant dimension's coefficients between two order groups. *Type* provides information on the type of order effect by comparing the effect sizes of the two coefficients with each other. "No order eff." indicates that the two coefficients are in the same position, so differences cannot be caused by the order of said dimension. We only compare those dimensions whose positions change at least once between the four order groups.
n.s.=not significant; + $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$

Appendix D: Is there a Recency Effect?

Table 4.D1. Wald Tests for Recency Effects

		Wald Test F *(p)*		
	N	Career orientation	Work experience	Siblings
Full sample	2,030	0.07 (.939)	3.31 (.069)	3.32* (.037)
Under 60 years	1,615	0.01 (.992)	0.69 (.408)	3.59* (.028)
60 years and older	415	0.50 (.609)	5.43* (.021)	0.03 (.975)
Higher education	1,515	0.10 (.090)	4.31* (.038)	2.74 (.065)
Lower education	515	0.25 (.782)	0.03 (.866)	0.97 (.382)
Fast responders	1,014	0.08 (.925)	2.28 (.132)	2.50 (.084)
Slow responders	1,016	0.06 (.938)	1.30 (.255)	1.10 (.334)

Table 4.D2. Full Sample: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Recency Dimensions (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (ref: German)	0.039	(0.104)	0.043	(0.104)	0.039	(0.104)
Male (ref: female)	−0.010	(0.111)	0.000	(0.111)	−0.002	(0.111)
Place of birth (ref: Konstanz)						
same province	0.196	(0.138)	0.196	(0.137)	0.196	(0.138)
other German city	0.082	(0.133)	0.096	(0.133)	0.097	(0.133)
GPA (ref: very good)						
good	−0.688***	(0.135)	−0.679***	(0.136)	−0.685***	(0.136)
satisfactory	−1.439***	(0.145)	−1.419***	(0.144)	−1.424***	(0.145)
Income parents (ref: low)						
middle	−1.076***	(0.138)	−1.070***	(0.139)	−1.069***	(0.139)
high	−3.968***	(0.148)	−3.985***	(0.147)	−3.973***	(0.148)
Siblings (ref: none)						
one sibling	0.506***	(0.149)	0.676***	(0.126)	0.674***	(0.126)
two siblings	0.867***	(0.154)	0.914***	(0.136)	0.911***	(0.136)
Prev. work exp. (ref: job train.)						
voluntary social year	0.834***	(0.109)	0.703***	(0.123)	0.829***	(0.109)
Career orient. (ref: career)						
family & career	−0.190	(0.132)	−0.176	(0.133)	−0.143	(0.185)
family	−0.627***	(0.139)	−0.622***	(0.139)	−0.626**	(0.194)
Last position (ref: no)						
siblings (yes)	−0.092	(0.241)				
prev. work exp. (yes)			−0.443*	(0.199)		
career orient. (yes)					0.059	(0.210)
Interactions: Last position X						
one sibling	0.682*	(0.271)				
two siblings	0.191	(0.324)				
voluntary social year			0.464	(0.255)		
family & career					−0.078	(0.265)
family					0.008	(0.279)
Constant	7.914***	(0.208)	7.990***	(0.209)	7.846***	(0.232)
Observations	2030		2030		2030	
R^2	0.333		0.333		0.331	

Note: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table 4.D3. 60 Years and Older: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Relevant Recency Dimension (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (ref: German)	−0.092	(0.242)	−0.121	(0.241)	−0.104	(0.242)
Male (ref: female)	0.308	(0.248)	0.287	(0.249)	0.315	(0.250)
Place of birth (ref: Konstanz)						
same province	0.055	(0.294)	0.070	(0.291)	0.067	(0.290)
other German city	0.534	(0.285)	0.515	(0.277)	0.574*	(0.288)
GPA (ref: very good)						
good	−0.816*	(0.318)	−0.793*	(0.320)	−0.804*	(0.315)
satisfactory	−1.120***	(0.306)	−1.112***	(0.301)	−1.107***	(0.309)
Income parents (ref: low)						
middle	−1.374***	(0.302)	−1.320***	(0.304)	−1.369***	(0.301)
high	−4.228***	(0.334)	−4.281***	(0.331)	−4.238***	(0.336)
Siblings (ref: none)						
one sibling	0.688*	(0.312)	0.720**	(0.269)	0.719**	(0.264)
two siblings	0.778*	(0.372)	0.730*	(0.314)	0.760*	(0.321)
Prev. work exp. (ref: job train.)						
voluntary social year	0.945***	(0.254)	0.578	(0.295)	0.983***	(0.260)
Career orient. (ref: career)						
family & career	−0.558	(0.312)	−0.480	(0.313)	−0.252	(0.415)
family	−1.078***	(0.308)	−1.025**	(0.314)	−0.967*	(0.384)
Last position (ref: no)						
siblings (yes)	0.345	(0.486)				
prev. work exp. (yes)			−0.833	(0.428)		
career orient. (yes)					0.184	(0.509)
Interactions: Last position X						
one sibling	0.123	(0.581)				
two siblings	−0.003	(0.650)				
voluntary social year			1.344*	(0.577)		
family & career					−0.608	(0.624)
family					−0.225	(0.618)
Constant	7.866***	(0.560)	8.200***	(0.559)	7.836***	(0.594)
Observations	415		415		415	
R^2	0.348		0.354		0.347	

Note: Controlled for education. Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.D4. Under 60 Years: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Relevant Recency Dimension (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (ref: German)	0.073	(0.118)	0.076	(0.118)	0.073	(0.118)
Male (ref: female)	−0.091	(0.124)	−0.079	(0.124)	−0.081	(0.125)
Place of birth (ref: Konstanz)						
same province	0.232	(0.157)	0.230	(0.156)	0.227	(0.157)
other German city	−0.024	(0.151)	−0.009	(0.151)	−0.012	(0.152)
GPA (ref: very good)						
good	−0.663***	(0.151)	−0.647***	(0.152)	−0.652***	(0.152)
satisfactory	−1.507***	(0.166)	−1.478***	(0.165)	−1.480***	(0.166)
Income parents (ref: low)						
middle	−0.994***	(0.157)	−0.992***	(0.157)	−0.990***	(0.157)
high	−3.902***	(0.165)	−3.918***	(0.165)	−3.913***	(0.165)
Siblings (ref: none)						
one sibling	0.483**	(0.168)	0.685***	(0.143)	0.682***	(0.143)
two siblings	0.893***	(0.170)	0.963***	(0.153)	0.957***	(0.152)
Prev. work exp. (ref: job train.)						
voluntary social year	0.829***	(0.122)	0.762***	(0.137)	0.824***	(0.122)
career orient. (ref: career)						
family & career	−0.119	(0.148)	−0.111	(0.148)	−0.102	(0.208)
family	−0.551***	(0.158)	−0.543***	(0.158)	−0.549*	(0.223)
Last position (ref: no)						
siblings (yes)	−0.188	(0.274)				
prev. work exp. (yes)			−0.322	(0.226)		
career orient. (yes)					0.047	(0.233)
Interactions: Last position X						
one sibling	0.809**	(0.308)				
two siblings	0.269	(0.372)				
voluntary social year			0.235	(0.283)		
family & career					−0.019	(0.296)
family					0.018	(0.316)
Constant	7.729***	(0.284)	7.739***	(0.284)	7.633***	(0.306)
Observations	1615		1615		1615	
R^2	0.336		0.334		0.333	

Note: Controlled for education. Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.D5. Higher Education Level: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Relevant Dimension (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (ref: German)	0.242*	(0.117)	0.237*	(0.117)	0.235*	(0.117)
Male (ref: female)	−0.032	(0.127)	−0.018	(0.127)	−0.018	(0.127)
Place of birth (ref: Konstanz)						
same province	0.334*	(0.156)	0.338*	(0.155)	0.345*	(0.156)
other German city	0.119	(0.153)	0.127	(0.153)	0.131	(0.154)
GPA (ref: very good)						
good	−0.742***	(0.151)	−0.719***	(0.151)	−0.728***	(0.152)
satisfactory	−1.552***	(0.171)	−1.518***	(0.170)	−1.527***	(0.171)
Income parents (ref: low)						
middle	−1.126***	(0.152)	−1.110***	(0.153)	−1.118***	(0.153)
high	−3.965***	(0.163)	−3.978***	(0.163)	−3.970***	(0.163)
Siblings (ref: none)						
one sibling	0.440*	(0.170)	0.611***	(0.146)	0.615***	(0.145)
two siblings	0.857***	(0.171)	0.894***	(0.153)	0.891***	(0.153)
Prev. work exp. (ref: job train.)						
voluntary social year	0.956***	(0.124)	0.783***	(0.138)	0.951***	(0.125)
Career orient. (ref: career)						
family & career	−0.259	(0.150)	−0.252	(0.149)	−0.199	(0.207)
family	−0.781***	(0.161)	−0.781***	(0.161)	−0.721**	(0.231)
Last position (ref: no)						
siblings (yes)	−0.240	(0.283)				
prev. work exp. (yes)			−0.384	(0.231)		
career orient. (yes)					0.113	(0.236)
Interactions: Last position X						
one sibling	0.713*	(0.316)				
two siblings	0.143	(0.374)				
voluntary social year			0.619*	(0.298)		
family & career					−0.119	(0.298)
family					−0.117	(0.323)
Constant	7.921***	(0.243)	7.939***	(0.246)	7.784***	(0.272)
Observations	1515		1515		1515	
R^2	0.356		0.356		0.354	

Note: Controlled for age. Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.D6. Lower Education Level: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Relevant Dimension (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (German)	−0.511*	(0.227)	−0.493*	(0.225)	−0.489*	(0.225)
Male (ref: female)	0.095	(0.232)	0.107	(0.234)	0.082	(0.236)
Place of birth (ref: Konstanz)						
same province	−0.285	(0.289)	−0.269	(0.287)	−0.323	(0.291)
other German city	−0.010	(0.268)	0.044	(0.263)	−0.005	(0.265)
GPA (ref: very good)						
good	−0.527	(0.301)	−0.545	(0.303)	−0.595*	(0.298)
satisfactory	−1.208***	(0.274)	−1.181***	(0.279)	−1.227***	(0.283)
Income parents (ref: low)						
middle	−1.047**	(0.318)	−1.060***	(0.315)	−1.031**	(0.317)
high	−4.048***	(0.332)	−4.045***	(0.330)	−4.014***	(0.336)
Siblings (ref: none)						
one sibling	0.672*	(0.301)	0.858**	(0.258)	0.847**	(0.260)
two siblings	0.935**	(0.334)	1.003***	(0.295)	0.994***	(0.293)
Prev. work exp. (ref: job train.)						
voluntary social year	0.541*	(0.220)	0.513	(0.263)	0.537*	(0.220)
Career orient. (ref: career)						
family & career	0.045	(0.288)	0.061	(0.291)	−0.021	(0.393)
family	−0.184	(0.276)	−0.180	(0.276)	−0.385	(0.356)
Last position (ref: no)						
siblings (yes)	0.366	(0.455)				
prev. work exp. (yes)			−0.666	(0.400)		
career orient. (yes)					−0.166	(0.452)
Interactions: Last position X						
one sibling	0.761	(0.559)				
two siblings	0.237	(0.637)				
voluntary social year			0.081	(0.482)		
family & career					0.146	(0.573)
family					0.390	(0.567)
Constant	7.797***	(0.437)	8.028***	(0.429)	7.986***	(0.485)
Observations	515		515		515	
R^2	0.305		0.303		0.297	

Note: Controlled for age. Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.D7. Fast Responders: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Relevant Dimension (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (ref: German)	0.075	(0.147)	0.078	(0.147)	0.070	(0.147)
Male (ref: female)	0.007	(0.159)	0.017	(0.159)	0.009	(0.159)
Place of birth (ref: Konstanz)						
same province	−0.013	(0.200)	−0.025	(0.199)	−0.021	(0.200)
other German city	−0.037	(0.183)	−0.032	(0.184)	−0.026	(0.184)
GPA (ref: very good)						
good	−0.580**	(0.194)	−0.573**	(0.195)	−0.585**	(0.195)
satisfactory	−1.392***	(0.213)	−1.378***	(0.210)	−1.372***	(0.213)
Income parents (ref: low)						
middle	−0.974***	(0.187)	−0.955***	(0.187)	−0.966***	(0.187)
high	−3.876***	(0.218)	−3.897***	(0.218)	−3.885***	(0.219)
Siblings (ref: none)						
one sibling	0.621**	(0.219)	0.825***	(0.182)	0.818***	(0.181)
two siblings	0.857***	(0.228)	0.872***	(0.193)	0.866***	(0.193)
Prev. work exp. (ref: job train.)						
voluntary social year	0.899***	(0.160)	0.750***	(0.181)	0.906***	(0.160)
Career orient. (ref: career)						
family & career	−0.202	(0.186)	−0.183	(0.186)	−0.178	(0.248)
family	−0.652**	(0.206)	−0.645**	(0.205)	−0.701*	(0.275)
Last position (ref: no)						
siblings (yes)	−0.089	(0.322)				
prev. work exp. (yes)			−0.423	(0.285)		
career orient. (yes)					−0.024	(0.299)
Interactions: Last position X						
one sibling	0.774*	(0.367)				
two siblings	0.019	(0.411)				
voluntary social year			0.543	(0.359)		
family & career					−0.028	(0.366)
family					0.115	(0.404)
Constant	7.948***	(0.307)	8.014***	(0.307)	7.921***	(0.331)
Observations	1014		1014		1014	
R^2	0.330		0.329		0.327	

Note: Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.D8. Slow Responders: OLS-Regression of Vignette Evaluation on Vignette Dimensions by Relevant Dimension (Siblings, Work Experience or Career Orientation)

	Siblings		Work exp.		Career orient.	
Foreign (ref: German)	−0.001	(0.151)	0.002	(0.149)	−0.000	(0.150)
Male (ref: female)	−0.013	(0.158)	−0.000	(0.158)	−0.002	(0.158)
Place of birth (ref: Konstanz)						
same province	0.427*	(0.190)	0.443*	(0.187)	0.440*	(0.189)
other German city	0.205	(0.195)	0.227	(0.194)	0.229	(0.196)
GPA (ref: very good)						
good	−0.802***	(0.191)	−0.798***	(0.191)	−0.802***	(0.192)
satisfactory	−1.484***	(0.203)	−1.461***	(0.203)	−1.479***	(0.204)
Income parents (ref: low)						
middle	−1.165***	(0.204)	−1.171***	(0.204)	−1.162***	(0.205)
high	−4.040***	(0.202)	−4.059***	(0.200)	−4.050***	(0.200)
Siblings (ref: none)						
one sibling	0.384	(0.204)	0.525**	(0.179)	0.526**	(0.178)
two siblings	0.875***	(0.208)	0.951***	(0.193)	0.948***	(0.193)
Prev. work exp. (ref: job train.)						
voluntary social year	0.746***	(0.149)	0.621***	(0.169)	0.730***	(0.151)
Career orient. (ref: career)						
family & career	−0.187	(0.192)	−0.179	(0.193)	−0.114	(0.278)
family	−0.581**	(0.191)	−0.583**	(0.191)	−0.547*	(0.275)
Last position (ref: no)						
siblings (yes)	−0.109	(0.360)				
prev. work exp. (yes)			−0.495	(0.281)		
career orient. (yes)					0.158	(0.293)
Interactions: Last position X						
one sibling	0.596	(0.402)				
two siblings	0.349	(0.498)				
voluntary social year			0.408	(0.358)		
family & career					−0.136	(0.382)
family					−0.062	(0.385)
Constant	7.869***	(0.287)	7.962***	(0.288)	7.754***	(0.326)
Observations	1016		1016		1016	
R^2	0.343		0.343		0.341	

Note: Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix E: Detailed Interpretation

Information on D-Efficiency

D-efficiency minimizes the correlations between dimensions (i.e., orthogonality) and maximizes the variance of each of the dimensions within the questionnaire versions (i.e., ensuring that each category occurs with about equal frequency, i.e., level balance). The optimization of both criteria (orthogonality and level balance) helps to enhance the precision with which one can estimate the parameters in statistical analyses (Auspurg and Hinz, 2015, p. 28; Kuhfeld et al., 1994; Kuhfeld, 2010). *D*-efficiency is equivalent to obtaining the most precise parameter estimates within OLS regressions, provided that the respondents produce valid judgments (Auspurg and Hinz, 2015, p. 29). *D*-efficiency has a range of [0; 100], with 100 indicating the most efficient designs (Kuhfeld, 2010; Kuhfeld et al., 1994). According to Auspurg and Hinz (2015, p. 29) designs with *D*-efficiency values over 90 offer sufficient statistical power to fulfill most research aims in the social sciences. For more information on *D*-efficiency see Auspurg and Hinz (2015), Kuhfeld et al. (1994) and Kuhfeld (2010).

Further Information on the Analysis Sample

The dimension work experience has three levels, with one of the levels being no information provided (see Table 4.A1 in Appendix A), which means that for a third of the vignettes the dimension was blank, i.e., left out of the description. This was done to study a possible statistical discrimination by gender. When this dimension is blank, the positions of some other dimensions change, e.g., in split 2 career orientation moves to the last position (and is no longer the penultimate). To allow for reliable estimates based on stable positions, we exclude all vignettes with blank information on work experience, starting with the test for overall order effects (second step). This reduces the number of vignettes to 2,030.

Interpretation of the Individual Wald Tests for the Subgroups of Age, Education, and Response Time

In both age subgroups, three out of 66 possible dimensions (4.5 percent) show statistically significant order effects. Likewise, in subgroup lower education level it is 1.5 percent, which is comparable to 3 percent in subgroup high education level. While for slow responders two dimensions show statistically significant order effects (3 percent), for fast responders it is five (7.6 percent) and therefore slightly more than could be expected to occur by chance. For tables see Tables 4.C2 - 4.C7 in Appendix C.

Recency: Interpretation of the Estimations for the Subgroups of Age, Education, and Response Time

As one might expect, Figure 4.2 (see section 4.4.2) shows that older people show statistically significant recency effects in the dimension work experience ($p = .021$), however, they do not show them in the dimensions siblings or career orientation. At the same time, younger people show statistically significant recency effects in one level of the dimension siblings ("one sibling", $p = .009$), but no recency effects in the dimensions previous work experience and career orientation. In addition, respondents with higher education level are affected by recency effects in one level of the dimension siblings ("one sibling", $p = .025$) and in the dimension previous work experience ($p = .038$), while respondents with lower education level did not show any recency effects (for regression tables see Tables 4.D3 - 4.D8 in Appendix D). The results are supported by joint Wald tests (see Table 4.D1 in Appendix D).

Robustness Tests and Additional Analyses

In a first robustness test we re-estimated our analysis, but this time by using random intercept models. This leads to similar conclusions. A notable difference is that by using a random intercept model, we find a statistically significant joint Wald Test ($\text{Chi}^2(33) = 49.64$, $p = .036$) within the subgroup of fast responders. This supports our finding that fast responders are more prone to order effects. The moderator test still reveals no differences between fast and slow responders ($\text{Chi}^2(33) = 33.16$, $p = .460$).

To consider the argument by Cohen (1969, p. 367-69), that the power of tests of interactions in a factorial design is distinctly lower than that of the main effects, we additionally considered $\alpha = .10$ as the significance threshold. This leads to similar conclusions.

We also did not find any recency effects for the small group of respondents with a lower education level who were over the age of 60. Furthermore, we tested a stricter classification for education level (i.e. academic degree as criteria for higher education) and age (i.e. respondents over the age of 70 years). The only additional result gained is that respondents of 70 years and older are more prone to order effects.

So far, we have left out all vignettes with blank information on the dimension work experience. As stated above, omitting to do this leads to position changes for some vignette dimensions, e.g., in split 2 for a third of the vignettes career orientation is in the last position (and not the penultimate). This allows us to conduct further tests for possible order effects, holding most other factors (i.e. the experimental split) constant. When recency effects arise, career orientation should impact the vignette evaluation more when it is in the last position. We run OLS regression using only the vignettes of split 2, and interacting career orientation with an indicator variable of whether it was

in the penultimate or last position of the vignette. Consistent with prior results, joint Wald tests did not show any indications for recency effects, either for the full sample or for any subgroup.

References

Aassve, Arnstein, Giulia Fuochi, and Letizia Mencarini. 2014. "Desparate Housework: Relative Resources, Time Availability, Economic Dependency, and Gender Ideology Across Europe." *Journal of Family Issues* 35:1000–1022.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2018. "Analyzing Causal Mechanisms in Survey Experiments." *Political Analysis* 26:357–378.

Ajzen, Icek. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50:179–211.

Akerlof, George A. and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115:715–753.

Amato, Paul R. and Alan Booth. 1995. "Changes in Gender Role Attitudes and Perceived Marital Quality." *American Sociological Review* 60:58–66.

Arránz Becker, Oliver. 2013. "Effects of Similarity of Life Goals, Values, and Personality on Relationship Satisfaction and Stability: Findings From a Two-Wave Panel Study." *Personal Relationships* 20:443–461.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by O. Ashenfelter and A. Rees, pp. 3–33. Princeton, NJ: Princeton University Press.

Atzmüller, Christiane and Peter M. Steiner. 2010. "Experimental Vignette Studies in Survey Research." *Methodology* 6:128–138.

Auspurg, Katrin and Sabine Düval. 2022. "Probing on the Conditionality of Gender Norms: Is the Significance of Gender Overrated?" *Unpublished Article.*

Auspurg, Katrin, Corinna Frodermann, and Thomas Hinz. 2014. "Berufliche Umzugsentscheidungen in Partnerschaften. Eine experimentelle Prüfung von Verhandlungstheorie, Frame-Selektion und Low-Cost-These." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 66:21–50.

Auspurg, Katrin and Thomas Hinz. 2015. *Factorial Survey Experiments*. Thousand Oaks, CA: Sage.

Auspurg, Katrin, Thomas Hinz, and Stefan Liebig. 2009. "Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey." *Methoden, Daten, Analysen (mda)* 3:59–96.

Auspurg, Katrin, Thomas Hinz, and Carsten Sauer. 2017a. "Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments." *American Sociological Review* 82:179–210.

Auspurg, Katrin, Maria Iacovou, and Cheti Nicoletti. 2017b. "Housework Share Between Partners: Experimental Evidence on Gender-Specific Preferences." *Social Science Research* 66:118–139.

Auspurg, Katrin and Annette Jäckle. 2017. "First Equals Most Important? Order Effects in Vignette-Based Measurement." *Sociological Methods & Research* 46:490–539.

Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review* 80:1095–1111.

Ayidiya, Stephen A. and McKee J. McClendon. 1990. "Response Effects in Mail Surveys." *Public Opinion Quarterly* 54:229–247.

Barth, Alice and Miriam Trübner. 2018. "Structural Stability, Quantitative Change: A Latent Class Analysis Approach Towards Gender Role Attitudes in Germany." *Social Science Research* 72:183–193.

Bartley, Sharon J., Priscilla W. Blanton, and Jennifer L. Gilliard. 2005. "Husbands and Wives in Dual-Earner Marriages: Decision-Making, Gender Role Attitudes, Division of Household Labor, and Equity." *Marriage & Family Review* 37:69–94.

Bauer, Gerrit. 2016. "Gender Roles, Comparative Advantages and the Life Course: The Division of Domestic Labor in Same-Sex and Different-Sex Couples." *European Journal of Population* 32:99–128.

Baxter, Janeen, Belinda Hewitt, and Michele Haynes. 2008. "Life Course Transitions and Housework: Marriage, Parenthood, and Time on Housework." *Journal of Marriage and Family* 70:259–272.

Beck, Michael and Karl-Dieter Opp. 2001. "Der faktorielle Survey and die Messung von Normen." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 53:283–306.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." *The Economic Journal* 75:493–517.

Becker, Gary S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* 3:33–58.

Becker, Gary S. 1993. *A Treatise on the Family*. Cambridge, Massachusetts: Harvard University Press.

Beere, Carole A. 1990. *Gender Roles: A Handbook of Tests and Measures*. Westport, CT: Greenwood Press.

Beere, Carole A., Daniel W. King, Donald B. Beere, and Lynda A. King. 1984. "The Sex-Role Egalitarianism Scale: A Measure of Attitudes Toward Equality Between the Sexes." *Sex Roles* 10:563–576.

Behr, Dorothée, Michael Braun, Lars Kaczmirek, and Wolfgang Bandilla. 2013. "Testing the Validity of Gender Ideology Items by Implementing Probing Questions in Web Surveys." *Field Methods* 25:124–141.

Beniger, James R. and Jodi A. Gusek. 1995. "The Cognitive Revolution in Public Opinion and Communication Research." In *Public Opinion and the Communication of Consent*, edited by T. L. Glasser and Ch. T. Salmon. New York: Guilford.

Bernard, Jessie. 1972. *The Future of Marriage*. New York: World Publishing.

Besbris, Max and Shamus Khan. 2017. "Less Theory. More Description." *Sociological Theory* 35:147–153.

Best, Henning and Clemens Kroneberg. 2012. "Die Low-Cost-Hypothese." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64:535–561.

Bicchieri, Christina. 2008. "The Fragility of Fairness: An Experimental Investigation on the Conditional Status of Pro-Social Norms." *Philosophical Issues* 18:229–248.

Bicchieri, Cristina. 2010. "Norms, Preferences, and Conditional Behavior." *Politics, Philosophy & Economics* 9:297–313.

Bicchieri, Cristina and Alex K. Chavez. 2010. "Behaving as Expected: Public Information and Fairness Norms." *Journal of Behavioral Decision Making* 23:161–178.

Bicchieri, Cristina and Alex K. Chavez. 2013. "Norm Manipulation, Norm Evasion: Experimental Evidence." *Economics and Philosophy* 29:175–198.

Bicchieri, Cristina, Jan W. Lindemans, and Ting Jiang. 2014. "A Structured Approach to a Diagnostic of Collective Practices." *Frontiers in Psychology* 5:1–13.

Bittman, Michael, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. 2003. "When Does Gender Trump Money? Bargaining and Time in Household Work." *American Journal of Sociology* 109:186–214.

Blair, Sampson L. and Daniel T. Lichter. 1991. "Measuring the Division of Household Labor Gender Segregation of Housework Among American Couples." *Journal of Family Issues* 12:91–113.

Blood, Robert O. and Donald M. Wolfe. 1960. *Husbands and Wives: The Dynamics of Married Living*. Glencoe, Illinois: Free Press.

Bolzendahl, Catherine I. and Daniel J. Myers. 2004. "Feminist Attitudes and Support for Gender Equality: Opinion Change in Women and Men, 1974-1998." *Social Forces* 83:759–789.

Braun, Michael. 1998. "Gender Roles." In *Comparative Politics: The Problem of Equivalence*, edited by J. W. Van Deth, p. 111–134. London: Routledge.

Braun, Michael. 2008. "Using Egalitarian Items to Measure Men's and Women's Family Roles." *Sex Roles* 59:644–656.

Braun, Michael and Jacqueline Scott. 2009. "Gender-Role Egalitarianism—Is the Trend Reversal Real?" *International Journal of Public Opinion Research* 21:362–367.

Braun, Michael, Jacqueline Scott, and Duane F. Alwin. 1994. "Economic Necessity or Self-Actualization? Attitudes Toward Women's Labour-Force Participation in East and West Germany." *European Sociological Review* 10:29–47.

Braun, Norman and Axel Franzen. 1995. "Umweltverhalten und Rationalität." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47:231–248.

Brines, Julie. 1993. "The Exchange Value of Housework." *Rationality and Society* 5:302–340.

Brines, Julie. 1994. "Economic Dependency, Gender, and the Division of Labor at Home." *American Journal of Sociology* 100:652–688.

Brüderl, Josef, Sonja Drobnič, Karsten Hank, Bernhard Nauck, Franz J. Neyer, Sabine Walper, Philipp Alt, Christiane Bozoyan, Petra Buhr, Christine Finn, Madison Garrett, Henriette Greischel, Nicolai Gröpler, Kristin Hajek, Michel Herzig, Bernadette Huyer-May, Rüdiger Lenke, Lara Minkus, Bettina Müller, Timo Peter, Claudia Schmiedeberg, Philipp Schütze, Nina Schumann, Carolin Thönnissen, Martin Wetzel, and Barbara Wilhelm. 2019. "The German Family Panel (pairfam)." GESIS Data Archive, Cologne. ZA5678 Data file Version 10.0.0, doi: 10.4232/pairfam.5678.10.0.0.

Buchmann, Claudia, Thomas A. DiPrete, and Anne McDaniel. 2008. "Gender Inequalities in Education." *Annual Review of Sociology* 34:319–337.

Campbell, Andrea Louise. 2012. "Policy Makes Mass Politics." *Annual Review of Political Science* 15:333–351.

Carlson, Daniel L. and Jamie L. Lynch. 2013. "Housework: Cause and Consequence of Gender Ideology?" *Social Science Research* 42:1505–1518.

Carriero, Renzo and Lorenzo Todesco. 2017. "The Interplay Between Equity and Gender Ideology in Perceived Housework Fairness: Evidence From an Experimental Vignette Design." *Sociological Inquiry* 87:561–585.

Ceci, Stephen J. 1991. "How Much Does Schooling Influence General Intelligence and Its Cognitive Components? A Reassessment of the Evidence." *Developmental Psychology* 27:703–722.

Cohen, Jacob. 1969. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Coltrane, Scott. 2000. "Research on Household Labor: Modeling and Measuring the Social Embeddedness of Routine Family Work." *Journal of Marriage and Family* 62:1208–1233.

Cordero-Coma, Julia and Gøsta Esping-Andersen. 2018. "The Intergenerational Transmission of Gender Roles: Children's Contribution to Housework in Germany." *Journal of Marriage and Family* 80:1005–1019.

Correll, Shelley, Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112:1297–1339.

Cotter, David, Joan M. Hermsen, and Reeve Vanneman. 2011. "The End of the Gender Revolution? Gender Role Attitudes from 1977 to 2008." *American Journal of Sociology* 117:259–89.

Daminger, Allison. 2019. "The Cognitive Dimension of Household Labor." *American Sociological Review* 84:609–633.

Davis, Shannon N. and Theodore N. Greenstein. 2009. "Gender Ideology: Components, Predictors, and Consequences." *Annual Review of Sociology* 35:87–105.

Dean, Liz, Brendan Churchill, and Leah Ruppanner. 2022. "The Mental Load: Building a Deeper Theoretical Understanding of How Cognitive and Emotional Labor Overload Women and Mothers." *Community, Work & Family* 25:13–29.

Dechant, Anna, Harald Rost, and Florian Schulz. 2014. "Die Veränderung der Hausarbeitsteilung in Paarbeziehungen. Ein Überblick über die Längsschnittforschung und neue empirische Befunde auf Basis der pairfam-Daten." *Zeitschrift für Familienforschung* 26:144–168.

Deutsch, Morton. 1975. "Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice?" *Journal of Social Issues* 31:137–149.

Diefenbach, Heike and Karl-Dieter Opp. 2007. "When and Why Do People Think There Should Be a Divorce?" *Rationality and Society* 19:485–517.

Diekmann, Andreas and Peter Preisendörfer. 1998. "Umweltbewusstsein und Umweltverhalten in Low- und High-Cost-Situationen: Eine empirische Überprüfung der Low-Cost-Hypothese." *Zeitschrift für Soziologie* 27:438–453.

Diekmann, Andreas and Peter Preisendörfer. 2003. "Green and Greenback: The Behavioral Effects of Environmental Attitudes in Low-Cost and High-Cost Situations." *Rationality and Society* 15:441–472.

Dotti Sani, Giulia Maria. 2014. "Men's Employment Hours and Time on Domestic Chores in European Countries." *Journal of Family Issues* 35:1023–1047.

Düval, Sabine. 2023. "Do Men and Women Really Have Different Gender Role Attitudes? Experimental Insight on Gender-Specific Attitudes toward Paid and Unpaid Work in Germany." *Social Science Research* 112:1–20.

Düval, Sabine and Katrin Auspurg. 2018. "Gender Ideologies or Simply a Measurement of Respondents' Beliefs About Inequalities?" Presentation at "Seminar Analytische Soziologie: Theorie und empirische Anwendungen", Venice International University, Venice, Italy.

Düval, Sabine and Katrin Auspurg. 2019. "Should Women Do More Housework? Experimental Insight on the Validity of Classical Measurements of Gender Norms." Presentation at ASA Annual Meeting 2019, New York.

Düval, Sabine and Katrin Auspurg. 2020. "The Factorial Survey Experiment on Distribution of Work in Partnerships in the German Family Panel (pairfam)." *pairfam Technical Paper* 14.

Düval, Sabine and Katrin Auspurg. 2022. "The Factorial Survey Experiment on "Distribution of Work in Partnerships" in the German Family Panel (pairfam) - Content and Technical Information on the Partner Vignettes." *pairfam Technical Paper* 20.

Düval, Sabine and Thomas Hinz. 2020. "Different Order, Different Results? The Effects of Dimension Order in Factorial Survey Experiments." *Field Methods* 32:23–37.

Elster, Jon. 1989. *The Cement of Society: A Study of Social Order*. Cambridge: Cambridge University Press.

Elster, Jon. 2011. "Norms." In *The Oxford Handbook of Analytical Sociology*, edited by Peter Bearman and Peter Hedström, Oxford Handbooks. Oxford University Press.

Elster, Jon. 2018. "Norms." In *The Oxford Handbook of Analytical Sociology*, edited by Peter Bearman and Peter Hedström. Oxford: Oxford University Press.

England, Paula. 2011. "Missing the Big Picture and Making Much Ado About Almost Nothing: Recent Scholarship on Gender and Household Work." *Journal of Family Theory & Review* 3:23–26.

Eurostat. 2022. "Gender Pension Gap by Age Group – EU-SILC Survey." Retrieved Novemer 01, 2022 (`https://ec.europa.eu/eurostat/databrowser/view/ilc_pnp13/default/table?lang=en`).

Eyerund, Theresa and Anja Karin Orth. 2019. "Geschlechterrollen: In der Theorie modern, in der Praxis klassisch." Report, German Economic Institute (IW).

Fortin, Nicole M. 2005. "Gender Role Attitudes and the Labour-Market Outcomes of Women across OECD Countries." *Oxford Review of Economic Policy* 21:416–438.

Fuwa, Makiko. 2004. "Macro-level Gender Inequality and the Division of Household Labor in 22 Countries." *American Sociological Review* 69:751–767.

Fuwa, Makiko and Philip N. Cohen. 2007. "Housework and Social Policy." *Social Science Research* 36:512–530.

Gager, Constance T. 1998. "The Role of Valued Outcomes, Justifications, and Comparison Referents in Perceptions of Fairness Among Dual-Earner Couples." *Journal of Family Issues* 19:622–648.

Gangl, Markus and Andrea Ziefle. 2015. "The Making of a Good Woman: Extended Parental Leave Entitlements and Mothers' Work Commitment in Germany." *American Journal of Sociology* 121:511–563.

Gartner, Hermann and Thomas Hinz. 2009. "Geschlechtsspezifische Lohnungleichheit in Betrieben, Berufen und Jobzellen (1993–2006)." *Berliner Journal für Soziologie* 19:557–575.

German Federal Statistical Office. 2015. "Zeitverwendungserhebung. Aktivitäten in Stunden und Minuten für Ausgewählte Personengruppen. 2012/2013." Retrieved November 01, 2022 (`https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Zeitverwendung/Publikationen/Downloads-Zeitverwendung/zeitverwendung-5639102139004.pdf?__blob=publicationFile`).

German Federal Statistical Office. 2022a. "Gender Pay Gap 2020: Germany Remains Among EU States with Highest Disparity." Retrieved November 01, 2022 (`https:`

//www.destatis.de/Europa/EN/Topic/Population-Labour-Social-Issues/Labour-market/gender_pay_gap.html).

German Federal Statistical Office. 2022b. "Unbereinigter Gender Pay Gap (GPG) nach Gebietsstand ab 1995." Retrieved November 01, 2022 (https://www.destatis.de/DE/Themen/Arbeit/Verdienste/Verdienste-Verdienstunterschiede/Tabellen/ugpg-01-gebietsstand.html;jsessionid=21954C53A8B64D8466C8D23376D44ED5.live722?nn=206824).

GESIS - Leibniz-Institut für Sozialwissenschaften. 2021. *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS - Kumulation 1980-2018*. GESIS Datenarchiv, Köln. ZA5274 Datenfile Version 1.1.0, https://doi.org/10.4232/1.13748.

Greenstein, Theodore N. 1996. "Gender Ideology and Perceptions of the Fairness of the Division of Household Labor: Effects on Marital Quality." *Social Forces* 74:1029–1042.

Gregory, Abigail and Susan Milner. 2011. "What is "New" about Fatherhood?: The Social Construction of Fatherhood in France and the UK." *Men and Masculinities* 14:588–606.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. "Survey Methodology." In *Wiley Series in Survey Methodology.*, edited by Robert M. Groves. Hoboken, NJ: J. Wiley.

Grunow, Daniela. 2019. "Comparative Analyses of Housework and Its Relation to Paid Work: Institutional Contexts and Individual Agency." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 71:247–284.

Grunow, Daniela, Katia Begall, and Sandra Buchler. 2018. "Gender Ideologies in Europe: A Multidimensional Framework." *Journal of Marriage and Family* 80:42–60.

Grunow, Daniela, Florian Schulz, and Hans-Peter Blossfeld. 2007. "Was erklärt die Traditionalisierungsprozesse häuslicher Arbeitsteilung im Eheverlauf: soziale Normen oder ökonomische Ressourcen?/What Explains the Process of Traditionalization in the Division of Household Labor: Social Norms or Economic Resources?" *Zeitschrift für Soziologie* 36:162–181.

Grunow, Daniela, Florian Schulz, and Hans-Peter Blossfeld. 2012. "What Determines Change in the Division of Housework Over the Course of Marriage?" *International Sociology* 27:289–307.

Gupta, Sanjiv. 1999. "The Effects of Transitions in Marital Status on Men's Performance of Housework." *Journal of Marriage and the Family* pp. 700–711.

Gupta, Sanjiv, Liana C. Sayer, and Jessica Pearlman. 2021. "Educational and Type of Day Differences in Mothers' Time Availability for Child Care and Housework." *Journal of Marriage and Family* 83:786–802.

Gustafson, Per and Urban Fransson. 2015. "Age Differences Between Spouses: Sociodemographic Variation and Selection." *Marriage & Family Review* 51:610–632.

Halimi, Myriam, Els Consuegra, Katrien Struyven, and Nadine Engels. 2018. "A Critical Examination of the Reliability and Validity of a Gender Role Attitude Scale in Flanders (Belgium): What Lessons Can Be Learned?" *Sex Roles* 78:423–438.

Hechter, Michael and Karl-Dieter Opp. 2001. "What Have We Learned About the Emergence of Social Norms?" In *Social Norms*, edited by M. Hechter and K.-D. Opp, pp. 394–415. New York: Russell Sage Foundation.

Hedström, Peter and Peter Bearman. 2009. *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.

Heisig, Jan Paul. 2011. "Who Does More Housework: Rich or Poor? A Comparison of 33 Countries." *American Sociological Review* 76:74–99.

Hinz, Thomas, Anja Joos, and Katharina Schleicher. 2012. "Bildung und Lebenszufriedenheit: Ergebnisse der Konstanzer Bürgerbefragung; 2011 - 4. Welle."

Hochschild, Arlie Russell and Anne Machung. 1989. *The Second Shift*. New York: Viking.

Holbrook, Allyson L., Jon A. Krosnick, David Moore, and Roger Tourangeau. 2007. "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes." *Public Opinion Quarterly* 71:325–348.

Hole, Arne. 2007. "WTP: Stata Module to Estimate Confidence Intervals for Willingness to Pay Measures."

Horne, Christine. 2003. "The Internal Enforcement of Norms." *European Sociological Review* 19:335–343.

Horne, Rebecca M. and Matthew D. Johnson. 2018. "Gender Role Attitudes, Relationship Efficacy, and Self-Disclosure in Intimate Relationships." *The Journal of Social Psychology* 158:37–50.

Hox, Joop J., Ita G. G. Kreft, and Piet L. J. Hermkens. 1991. "The Analysis of Factorial Surveys." *Sociological Methods & Research* 19:493–510.

Huinink, Johannes, Josef Brüderl, Bernhard Nauck, Sabine Walper, Laura Castiglioni, and Michael Feldhaus. 2011. "Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual Framework and Design." *Zeitschrift für Familienforschung* 23:77–101.

ISSP. 2021. "Family and Changing Gender Roles V." Source Questionnaire. ISSP 2022. Retrieved Novemer 01, 2022 (`https://issp.org/wp-content/uploads/2022/02/issp_2022_final_source_questionnaire.pdf`).

Jacobs, Jerry A. and Kathleen Gerson. 2016. "Unpacking Americans' Views of the Employment of Mothers and Fathers Using National Vignette Survey Data: SWS Presidential Address." *Gender & Society* 30:413–441.

Jann, Ben. 2014. "Plotting Regression Coefficients and Other Estimates." *Stata Journal* 14:708–737.

Jasso, Guillermina. 1983. "Fairness of Individual Rewards and Fairness of the Reward Distribution: Specifying the Inconsistency Between the Micro and Macro Principles of Justice." *Social Psychology Quarterly* 46:185–199.

Jasso, Guillermina and Karl-Dieter Opp. 1997. "Probing the Character of Norms: A Factorial Survey Analysis of the Norms of Political Action." *American Sociological Review* 62:947–964.

Johnston, Deirdre D. and Debra H. Swanson. 2006. "Constructing the "Good Mother": The Experience of Mothering Ideologies by Work Status." *Sex Roles* 54:509–519.

Kahnemann, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint of Profit Seeking: Entitlements in the Market." *The American Economic Review* 76:728–741.

Katrňák, Tomáš and Beatrice Chromková Manea. 2020. "Change in Prevalence or Preference? Trends in Educational Homogamy in Six European Countries in a Time of Educational Expansion." *Social Science Research* 91:102460.

Katz-Wise, Sabra L., Heather A. Priess, and Janet S. Hyde. 2010. "Gender-Role Attitudes and Behavior Across the Transition to Parenthood." *Developmental Psychology* 46:18–28.

Kaufman, Gayle and Eva Bernhardt. 2015. "Gender, Work and Childbearing: Couple Analysis of Work Adjustments After the Transition to Parenthood." *Community, Work & Family* 18:1–18.

Kaufman, Gayle and Hiromi Taniguchi. 2019. "Gender Equality and Work–Family Conflict From a Cross-National Perspective." *International Journal of Comparative Sociology* 60:385–408.

Kelly, Maura and Elizabeth Hauck. 2015. "Doing Housework, Redoing Gender: Queer Couples Negotiate the Household Division of Labor." *Journal of GLBT Family Studies* 11:438–464.

Khoudja, Yassine and Fenella Fleischmann. 2018. "Gender Ideology and Women's Labor Market Transitions Within Couples in the Netherlands." *Journal of Marriage and Family* 80:1087–1106.

Killewald, Alexandra and Margaret Gough. 2010. "Money isn't Everything: Wives' Earnings and Housework Time." *Social Science Research* 39:987–1003.

Klein, Markus and Michael Kühhirt. 2010. "Sozial Erwünschtes Antwortverhalten Bezüglich der Teilung Häuslicher Arbeit: die Interaktion von Interviewergeschlecht und Befragtenmerkmalen in Telefoninterviews." *Methoden, Daten, Analysen (mda)* 4:79–104.

Knäuper, Bärbel. 1999. "The Impact of Age and Education on Response Order Effects in Attitude Measurement." *Public Opinion Quarterly* 63:347–370.

Knight, Carly R. and Mary C. Brinton. 2017. "One Egalitarianism or Several? Two Decades of Gender-Role Attitude Change in Europe." *American Journal of Sociology* 122:1485–1532.

Kroska, Amy. 2000. "Conceptualizing and Measuring Gender Ideology as an Identity." *Gender & Society* 14:368–394.

Kroska, Amy. 2007. "Gender Ideology and Gender Role Ideology." In *The Blackwell Encyclopedia of Sociology*, edited by G. Ritzer. John Wiley & Sons.

Kroska, Amy and Cheryl Elman. 2009. "Change in Attitudes About Employed Mothers: Exposure, Interests, and Gender Ideology Discrepancies." *Social Science Research* 38:366–382.

Krosnick, Jon A. 1989. "Attitude Importance and Attitude Accessibility." *Personality and Social Psychology Bulletin* 15:297–308.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213–236.

Krosnick, Jon A. and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *The Public Opinion Quarterly* 51:201–219.

Kuhfeld, Warren F. 2010. *Marketing Research Methods in SAS. Experimental Design, Choice, Conjoint and Graphical Techniques*. Cary, NC, USA: Cary.

Kuhfeld, Warren F., Randall D. Tobias, and Mark Garratt. 1994. "Efficient Experimental Design with Marketing Research Applications." *Journal of Marketing Research* 31:545–557.

Kühhirt, Michael. 2012. "Childbirth and the Long-Term Division of Labour Within Couples: How Do Substitution, Bargaining Power, and Norms Affect Parents' Time Allocation in West Germany?" *European Sociological Review* 28:565–582.

Lachance-Grzela, Myléne and Geneviéve Bouchard. 2010. "Why Do Women Do the Lion's Share of Housework? A Decade of Research." *Sex Roles* 63:767–780.

Leopold, Thomas and Jan Skopek. 2015. "Convergence or Continuity? The Gender Gap in Household Labor After Retirement." *Journal of Marriage and Family* 77:819–832.

Mayerl, Jochen, Piet Selke, and Dieter Urban. 2005. "Analyzing Cognitive Processes in CATI-Surveys with Response Latencies: An Empirical Evaluation of the Consequences of Using Different Baseline Speed Measures." *Schriftreihe des Instituts für Sozialwissenschaften der Universität Stuttgart* 2:1–18.

Mayerl, Jochen and Dieter Urban. 2008. *Antwortreaktionszeiten in Survey-Analysen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Mays, Anja. 2012. "Determinanten Traditionell-Sexistischer Einstellungen in Deutschland – eine Analyse mit Allbus-Daten." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 64:277–302.

McClendon, McKee J. 1991. "Acquiescence and Recency Response-Order Effects in Interview Surveys." *Sociological Methods & Research* 20:60–103.

McClintock, Elizabeth Aura. 2017. "Occupational Sex Composition and Gendered Housework Performance: Compensation or Conventionality?" *Journal of Marriage and Family* 79:475–510.

McGinnity, Frances and Patricia McManus. 2007. "Paying the Price for Reconciling Work and Family Life: Comparing the Wage Penalty for Women's Part-Time Work in Britain, Germany and the United States." *Journal of Comparative Policy Analysis: Research and Practice* 9:115–134.

McHugh, Maureen C. and Irene Hanson Frieze. 1997. "The Measurement of Gender-Role Attitudes. A Review and Commentary." *Psychology of Women Quarterly* 21:1–16.

Meagher, Kelsey D. and Xiaoling Shu. 2019. "Trends in U.S. Gender Attitudes, 1977 to 2018: Gender and Educational Disparities." *Socius* 5:1–3.

Mincer, Jacob. 1962. *Labor Force Participation of Married Women: A Study of Labor Supply*. Aspects of Labor Economics. National Bureau of Economic Research, Inc.

Minnotte, Krista Lynn, Michael Charles Minnotte, and Daphne E. Pedersen. 2013. "Marital Satisfaction Among Dual-Earner Couples: Gender Ideologies and Family-to-Work Conflict." *Family Relations* 62:686–698.

Minnotte, Krista Lynn, Michael C. Minnotte, Daphne E. Pedersen, Susan E. Mannon, and Gary Kiger. 2010. "His and Her Perspectives: Gender Ideology, Work-to-Family Conflict, and Marital Satisfaction." *Sex Roles* 63:425–438.

Murray, Susan B. 1996. "'We All Love Charles': Men in Child Care and the Social Construction of Gender." *Gender and Society* 10:368–85.

Mutz, Diana C. 2011. *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press.

Myers, Scott M. and Alan Booth. 2002. "Forerunners of Change in Nontraditional Gender Ideology." *Social Psychology Quarterly* 65:18–37.

Narayan, Sowmya and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurment." *Public Opinion Quarterly* 60:58–88.

Nitsche, Natalie and Daniela Grunow. 2016. "Housework Over the Course of Relationships: Gender Ideology, Resources, and the Division of Housework From a Growth Curve Perspective." *Advances in Life Course Research* 29:80–94.

Nordenmark, Mikael. 2004. "Does Gender Ideology Explain Differences Between Countries Regarding the Involvement of Women and of Men in Paid and Unpaid Work?" *International Journal of Social Welfare* 13:233–243.

OECD. 2017. *The Pursuit of Gender Equality. An Uphill Battle.* Paris: OECD Publishing.

Opp, Karl-Dieter. 2001. "Norms." In *International Encyclopedia of the Social & Behavioral Sciences*, edited by N. J. Smelser and P. B. Baltes. Elsevier Science Ltd.

Opp, Karl-Dieter. 2002. "When Do Norms Emerge by Human Design and When by the Unintended Consequences of Human Action? The Example of the No-smoking Norm." *Rationality and Society* 14:131–158.

Opp, Karl-Dieter. 2020. "Normen und Institutionen: Entstehung, Wandel und Wirkungen. Eine Anwendung der Theorie rationalen Handelns." In *Rational Choice*, edited by A. Tutić, pp. 232–259. De Gruyter Oldenbourg.

Ott, Notburga. 1992. *Intrafamily Bargaining and Household Decisions*. Berlin: Springer.

pairfam Group. 2019. "Anchor Codebook Wave 10 (2017/2018), Release 10.0." GESIS Data Archive, Cologne. ZA5678 Data file Version 10.0.0, doi: 10.4232/pairfam.5678.10.0.0.

Parsons, Talcott. 1955. "The American Family: Its Relations to Personality and to the Social Structure." In *Family, Socialization and Interaction Process*, edited by T. Parsons and R. F. Bales. New York: The Free Press. Macmillan Publishing.

Parsons, Talcott. 1959. "The Social Structure of the Family." In *The Family: Its Function and Destiny*, edited by R. N. Anshen, pp. 241–274. New York: Harper & Row.

Pedulla, David S. and Sarah Thébaud. 2015. "Can We Finish the Revolution? Gender, Work-family Ideals, and Institutional Constraint." *American Sociological Review* 80:116–139.

Perales, Francisco, Janeen Baxter, and Tsui-o Tai. 2015. "Gender, Justice and Work: A Distributive Approach to Perceptions of Housework Fairness." *Social Science Research* 51:51–63.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62:659–661.

Presser, Harriet B. 1994. "Employment Schedules Among Dual-Earner Spouses and the Division of Household Labor byGender." *American Sociological Review* 59:348–364.

Procher, Vivien, Nolan Ritter, and Colin Vance. 2018. "Housework Allocation in Germany: The Role of Income and Gender Identity." *Social Science Quarterly* 99:43–61.

Rauhut, Heiko and Ivar Krumpal. 2008. "Die Durchsetzung sozialer Normen in Low-Cost und High-Cost Situationen / Enforcement of Social Norms in Low-Cost and High-Cost Situations." *Zeitschrift für Soziologie* 37:380–402.

Ridgeway, Cecilia L. 2011. *Framed by Gender. How Gender Inequality Persists in the Modern World*. Oxford: Oxford University Press.

Robbins, Blaine and Edgar Kiser. 2018. "Legitimate Authorities and Rational Taxpayers: An Investigation of Voluntary Compliance and Method Effects in a Survey Experiment of Income Tax Evasion." *Rationality and Society* 30:247–301.

Rogers, William. 1994. "Regression Standard Errors in Clustered Samples." *Stata Technical Bulletin* 3.

Ruppanner, Leah, Ben Maltby, Belinda Hewitt, and David Maume. 2021. "Parents' Sleep Across Weekdays and Weekends: The Influence of Work, Housework, and Childcare Time." *Journal of Family Issues* Online First.

Sanchez, Laura and Elizabeth Thomson. 1997. "Becoming Mothers and Fathers: Parenthood, Gender, and the Division of Labor." *Gender & Society* 11:747–772.

Sauer, Carsten, Katrin Auspurg, and Thomas Hinz. 2020. "Designing Multi-Factorial Survey Experiments: Effects of Presentation Style (Text or Table), Answering Scales, and Vignette Order." *methods, data, analyses* 14:195–214.

Sauer, Carsten, Katrin Auspurg, Thomas Hinz, and Stefan Liebig. 2011. "The Application of Factorial Surveys in General Population Samples: The Effects of Respondent Age and Education on Response Times and Response Consistency." *Survey Research Methods* 5:89–102.

Scarborough, William J., Joanna R. Pepin, Danny L. Lambouths, Ronald Kwon, and Ronaldo Monasterio. 2021. "The Intersection of Racial and Gender Attitudes, 1977 through 2018." *American Sociological Review* 86:823–855.

Scarborough, William J. and Barbara J. Risman. 2017. "Changes in the Gender Structure: Inequality at the Individual, Interactional, and Macro Dimensions." *Sociology Compass* 11:1–16.

Schaie, K. Warner. 1996. *Intellectual Development in Adulthood*. Handbook of the Psychology of Aging, 4th ed. San Diego, CA, US: Academic Press.

Schneider, Daniel. 2011. "Market Earnings and Household Work: New Tests of Gender Performance Theory." *Journal of Marriage and Family* 73:845–860.

Schneider, Daniel. 2012. "Gender Deviance and Household Work: The Role of Occupation." *American Journal of Sociology* 117:1029–1072.

Schober, Pia S. 2013. "The Parenthood Effect on Gender Inequality: Explaining the Change in Paid and Domestic Work When British Couples Become Parents." *European Sociological Review* 29:74–85.

Schulz, Florian. 2021. "Attitudes Towards Sharing Housework in Couple Context: An Empirical, Factorial Survey Approach." *Journal of Family Research* 33:1–36.

Schulz, Florian and Hans-Peter Blossfeld. 2006. "Wie verändert sich die häusliche Arbeitsteilung im Eheverlauf?" *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58:23–49.

Schwarz, Norbert. 2007. "Cognitive Aspects of Survey Methodology." *Applied Cognitive Psychology* 21:277–287.

Schwarz, Norbert and Bärbel Knäuper. 2000. "Cognition, Aging, and Self-Reports." In *Cognitive Aging: A Primer*, edited by D. Park and N. Schwarz, pp. 233–52. Philadelphia: Psychology Press.

Schwarz, Norbert, Fritz Strack, Hans-J. Hippler, and George Bishop. 1991. "The Impact of Administration Mode on Response Effects in Survey Measurement." *Applied Cognitive Psychology* 5:193–212.

Scott Carter, J., Mamadi Corra, and Shannon K. Carter. 2009. "The Interaction of Race and Gender: Changing Gender-Role Attitudes, 1974–2006." *Social Science Quarterly* 90:196–211.

Solera, Cristina and Letizia Mencarini. 2018. "The Gender Division of Housework After the First Child: A Comparison Among Bulgaria, France and the Netherlands." *Community, Work & Family* 21:519–540.

StataCorp. 2019. *Stata Statistical Software: Release 16*. Texas: StataCorp LLC.

Stoet, Gijsbert and David C. Geary. 2018. "The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education." *Psychological Science* 29:581–593.

Sullivan, Oriel. 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." *Journal of Family Theory & Review* 3:1–13.

Sullivan, Oriel. 2021. "The Gender Division of Housework and Child Care." In *Research Handbook on the Sociology of the Family*, edited by N. F. Schneider and M. Kreyenfeld, p. 342–354. UK and Massachusetts, USA: Edward Elgar Publishing Limited.

Szinovacz, Maximiliane E. 2000. "Changes in Housework After Retirement: A Panel Analysis." *Journal of Marriage and Family* 62:78–92.

Tai, Tsui-o and Judith Treas. 2012. "Housework Task Hierarchies in 32 Countries." *European Sociological Review* 29:780–791.

Thébaud, Sarah. 2010. "Masculinity, Bargaining, and Breadwinning: Understanding Men's Housework in the Cultural Context of Paid Work." *Gender & Society* 24:330–354.

Thébaud, Sarah, Sabino Kornrich, and Leah Ruppanner. 2021. "Good Housekeeping, Great Expectations: Gender and Housework Norms." *Sociological Methods & Research* 50:1186–1214.

Thibaut, John W. and Harold H. Kelley. 1959. *The Social Psychology of Groups*. The Social Psychology of Groups. New York: Wiley.

Thiel, Fabian. 2020. "Die Low-Cost-Hypothese. Ein empirischer Test am Beispiel der Befürwortung einer City-Maut." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 72:429–453.

Thiel, Fabian, Sabine Düval, and Katrin Auspurg. 2021. "Robustness of Mulitfactorial Survey Experiments to Explore Environmental Attitudes. Probing on the Effects of Presentation Format, Number of Levels, and Dimension Order." Presentation at 2nd Meeting on Environmental Social Sciences, Nuffield College, University of Oxford, UK.

Thompson, Linda. 1991. "Family Work: Women's Sense of Fairness." *Journal of Family Issues* 12:181–196.

Trappe, Heike, Matthias Pollmann-Schult, and Christian Schmitt. 2015. "The Rise and Decline of the Male Breadwinner Model: Institutional Underpinnings and Future Expectations." *European Sociological Review* 31:230–242.

Urban, Dieter and Jochen Mayerl. 2007. "Antwortlatenzzeiten in der Survey-Basierten Verhaltensforschung." *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59:692–713.

Van Vlimmeren, Eva, Guy B. D. Moors, and John P. T. M. Gelissen. 2017. "Clusters of Cultures: Diversity in Meaning of Family Value and Gender Role Items Across Europe." *Quality & Quantity* 51:2737–2760.

Wallander, Lisa. 2009. "25 Years of Factorial Surveys in Sociology: A Review." *Social Science Research* 38:505–520.

Walper, Sabine and Shih-cheng Lien. 2018. "Routinebetreuung und interaktive „Quality Time": Was beeinflusst, wieviel Zeit Väter wie mit ihren Kindern verbringen?/Routine Child Care and Interactive "Quality Time": What Affects How Much Time Fathers Spend With Their Children and How They Do This?" *Journal of Family Research* 30:29–49.

Walter, Jessica Gabriele. 2018a. "The Adequacy of Measures of Gender Roles Attitudes: A Review of Current Measures in Omnibus Surveys." *Quality & Quantity* 52:829–848.

Walter, Jessica Gabriele. 2018b. "Measures of Gender Role Attitudes under Revision: The Example of the German General Social Survey." *Social Science Research* 72:170–182.

West, Candace and Don H. Zimmerman. 1987. "Doing Dender." *Gender & Society* 1:125–151.

Wood, Wendy and Alice H. Eagly. 2015. "Two Traditions of Research on Gender Identity." *Sex Roles* 73:461–473.

Wooldridge, Jeffey M. 2009. *Introductory Econometrics. A Modern Approach*, volume 4. Mason, OH: South Western.

Yavorsky, Jill E., Claire M. Kamp Dush, and Sarah J. Schoppe-Sullivan. 2015. "The Production of Inequality: The Gender Division of Labor Across the Transition to Parenthood." *Journal of Marriage and Family* 77:662–679.

Yu, Wei-hsin and Pei-lin Lee. 2013. "Decomposing Gender Beliefs: Cross-National Differences in Attitudes Toward Maternal Employment and Gender Equality at Home." *Sociological Inquiry* 83:591–621.

Zuo, Jiping. 2004. "Shifting the Breadwinning Boundary:The Role of Men's Breadwinner Status and Their Gender Ideologies." *Journal of Family Issues* 25:811–832.