

IZA DP No. 8569

Housework Share between Partners: Experimental Evidence on Gender Identity

Katrin Auspurg
Maria Iacovou
Cheti Nicoletti

October 2014

Housework Share between Partners: Experimental Evidence on Gender Identity

Katrin Auspurg

Goethe University Frankfurt Main

Maria Iacovou

University of Cambridge

Cheti Nicoletti

*DERS, University of York,
ISER, University of Essex and IZA*

Discussion Paper No. 8569
October 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Housework Share between Partners: Experimental Evidence on Gender Identity^{*}

Using an experimental design, we investigate the reasons behind the gendered division of housework within couples. In particular, we assess whether the fact that women do more housework may be explained by differences in preferences deriving from differences in gender identity between men and women. We find little evidence of any systematic gender differences in the preference for housework, suggesting that the reasons for the gendered division of housework lie elsewhere.

JEL Classification: J16, J22, C35

Keywords: gender, housework, unpaid work, division of labor, experiment

Corresponding author:

Cheti Nicoletti
Department of Economics and Related Studies
University of Essex
Heslington, York YO10 5DD
United Kingdom
E-mail: cheti.nicoletti@york.ac.uk

^{*} This research work was supported by the Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MISOC) (award no. res-518-28-001) and it is based on a survey factorial experiment which was funded by the UK Longitudinal Studies Centre (award no. RES-586-47-0002). We are also thankful to the Development Fund of the University of Konstanz and University of Essex.

1. INTRODUCTION

In this paper we report on the results of a novel experiment to assess the role of gender identity in determining the level of utility which men and women derive from different allocations of housework between partners.

The question of why women spend more time doing housework than their male partners has attracted considerable interest from scholars from across the social sciences (Becker 1965; Oakley 1974; Hakim 1996 and 2000; Akerlof and Kranton 2000; Baker and Jacobsen 2007; Lachance-Grzela and Bouchard 2010; Stratton 2012; and many others). This question is all the more relevant in the contemporary context where women's labor market participation and earnings have increased vastly relative to those of men, and where women's qualifications are now on a par with those of men, but where there remain large disparities in the amount of housework done by men and women (Kan et al. 2011).

A range of theories have been advanced to explain this phenomenon. These will be discussed fully in Section 2, but the debate essentially boils down to whether women do more housework because their capabilities and characteristics are systematically different from those of men; or because they are responding to pressure arising from power dynamics within the partner relationship or from society at large; or because women's gender identity means that they actually prefer to spend a greater proportion of their time doing housework than men do. It is this issue of differences in preferences arising from gender identity which this paper sets out to explore.

The first model of economic behavior which incorporated gender identity into the utility function was proposed by Akerlof and Kranton (2000); this model has the potential to explain asymmetries in the allocation of paid and unpaid work which conventional utility-based models cannot. We test such a model in an experimental framework, using as a measure of utility the levels of self-reported satisfaction in a factorial survey experiment, and testing whether gender is a relevant determinant of the utility individuals derive from housework.

Economists are often skeptical about the use of self-reported measures of subjective wellbeing, satisfaction, etc. (see e.g. Bertrand and Mullainathan 2001), due to concerns that individuals might understand subjective questions differently, or might differ in their use of the scale provided to rank their situation. This heterogeneity in response style across

individuals may lead to biased estimates; however, this is less of an issue in our experimental context, for two reasons. First, in our experimental design the principal factors affecting utility are randomly assigned, therefore avoiding any potential correlation between these factors and unobserved individual characteristics. Second, we explicitly take account of the issue of heterogeneity by collecting multiple observations on participants and controlling for individual effects, showing that unobserved individual-specific characteristics (such as personality traits) are uncorrelated with the experimental factors explaining the level of satisfaction (utility).

There currently exists hardly any empirical evidence on how gender, and gender identity, affect the utility which individuals derive from the division of labor between partners. In fact, it is far from straightforward to use survey data to test empirically whether men and women have different preferences over the allocation of paid or unpaid work. There is no shortage of available data: several household surveys include questions on the allocation of paid work and housework between partners, and on individuals' satisfaction with these arrangements. However, these data contain only reports of people's satisfaction with their *actual* arrangements, and not the level of satisfaction that they *would* experience under alternative arrangements. This leads to three main problems. First, some distributions of housework and paid work are rarely observed in surveys (for example, households where the woman does more paid work, and earns more, and does less housework, than her male partner, are present only in very low numbers in most survey-based samples). This means that it is not possible to estimate preferences over the entire range of potential distributions of housework and paid work, because there are simply too few observations in some parts of the full space.

Second, people's satisfaction with the situation in which they actually find themselves may be affected by a process of ex-post rationalization, and may be a poor reflection of what their preferences might be, given the possibility of one or more alternative situations.

Third, and perhaps most importantly, people's actual hours of domestic and market work, as well as other factors such as their wage levels, are largely determined by their own characteristics and those of their partners – and some of these may be the same characteristics which drive people's utility with housework arrangements. Since surveys do not usually provide details on all potential relevant individual characteristics, empirical analyses are subject to problems with endogeneity, which means that it is difficult to draw causal inferences from survey data as to whether women's greater contribution to housework arises

as the result of gender-specific preferences, or as the result of a process of specialization triggered by partners' differences in productivity in the market and in the home.

In many contexts where behavior is endogenously determined, a randomized experiment would address the problem. However, the difficulties in carrying out a real-world randomized experiment in this context are obvious and insurmountable: it would simply not be possible to randomly allocate paid work, earnings or housework among a sample of couples.

An alternative empirical approach is the use of laboratory, field or survey experimental designs (see Croson and Gneezy 2009, Bertrand 2011). Three examples of experimental studies on gender identity are the lab experiment run by Cadsby et al. (2013) to test the effect of gender identity on attitudes to risk and competition, the lab experiment by Gorges (2014) to test gender specific patterns in couples' work specialization decisions and the factorial survey experiment adopted by Abraham et al. (2010) to test the effect of gender role attitudes on migration decisions within dual-earner partners. However, no experiment has previously been carried out to test differences in the perceived utility of housework arrangements between partners.

The current experiment is designed as follows. We invite people to imagine themselves and their partners in several different hypothetical domestic scenarios ("vignettes"), and to tell us how satisfied they would be with each set of arrangements. We generate these hypothetical scenarios using a multi-factorial experimental survey design¹; that is, as well as varying the distribution of housework between the different scenarios, we also vary a range of other factors: the share of paid work done by each partner; the level of respondents' own earnings and their partners' earnings; the presence and age of children in the home; and whether the household employs paid help (i.e. whether there is some market substitution of domestic work).

¹ Factorial survey experiments have been widely used by sociologists to study beliefs, attitudes and decisions (see Wallander 2009 for a review). Economists have used similar methods to study individual choice and willingness to pay, preferences across products for marketing purposes, evaluations of non-market goods such as health and environmental conditions, and to assess the utility of objects and situations ('stated preference experiments', 'stated choice experiments' and 'conjoint valuation methods' in e.g. Green and Srinivasan 1990; Louviere et al. 2000; Bateman et al. 2002; Amaya-Amaya et al. 2008; Sándor and Franses 2009).

The experimental design is described in detail in Section 3. A feature of this design is that vignettes are randomly allocated between households, with male and female members of couples receiving sets of vignettes which are identical but “reflected” (that is, the same housework and paid work arrangements, but with the roles of the male and female partners exchanged). This design allows us to assess directly whether the perceived utility derived from different work arrangements differs between men and women. A finding that there are systematic differences between men’s and women’s utility (and/or that both men and women prefer arrangements under which the woman does most of the housework) would lead us to conclude that gender identity, i.e. the internalization of social gender norms, affects the derived utility of men and women, as suggested by Akerlof and Kranton (2000). Conversely, finding that there are few or no differences in perceived utility between men and women (and the absence of a preference for gendered arrangements among either men or women) would lead us to conclude that gender identity does not play a significant role in determining the level of utility arising from different working arrangements; that in general, preferences over the division of work between partners are the same for men and women; and that the tendency of women to specialize more in housework is not due to preferences, but must be due to some other factor: women’s comparative advantage in domestic activities, as suggested by Becker (1965), or social gender norms that are not internalized by women, but that are nevertheless enforced.

The main analysis in this paper is carried out via linear random effects regressions; in addition, we carry out a range of validity checks and sensitivity analyses, including the estimation of a linear regression with fixed effects and an ordered probit specification with random effects, checks on whether unobserved characteristics such as mood and personality traits might affect the level of reported satisfaction; and checks on whether people abstract from their own gender when answering vignette questions. We also repeat our analysis on subsamples of individuals, in order to test whether gender differences, which are not evident across the population in general, may be present for groups in specific circumstances: (i) those who are actually married or living with a partner, (ii) those who have children in real life, and (iii) people who undertake a high share of housework in real life.

2. BACKGROUND AND LITERATURE

Becker (1965) models choices relating to the allocation of housework and paid work between partners as being determined by the returns from specialization: if one partner is relatively

more productive than the other in market work, the overall utility accruing to a household will be maximized if that partner specializes in market work while the other specializes in housework. This would explain why, in a context where men have higher levels of human capital than women, men do the majority of paid work and women do most of the housework. However, it also implies that if women and men had identical levels of human capital, we should observe both sexes doing similar shares of market work and housework. In fact, in real life, women generally do more housework than men, even when partners have similar levels of education (Brines 1993, 1994; Akerlof and Kranton, 2000).

Becker (1985) explains this gender asymmetry in housework shares as a consequence of the comparative advantage of women in childcare and in other domestic activities such as cooking and cleaning. Women's comparative advantage in these areas means that they end up doing more childcare and housework, and consequently choose paid jobs that are more flexible, less demanding and less well rewarded. This is the main explanation that Becker (1985) proposes for the gendered division of housework; however, he notes that gender asymmetry could also be justified by any other factor which leads women to have lower-paid jobs, including (but not limited to) discrimination against women, social norms, "gender exploitation", and work interruptions for childbearing. As long as women tend to be paid less than men in the labor market for whatever reason, they have a comparative advantage for specializing more in housework. In theory this specialization is gender-neutral, meaning that if a man were comparatively more productive than his female partner in the domestic sphere, he should do a higher share of the housework than his partner. However, in real life, women are observed to do a larger share of housework even when their market work share is as large as, or even larger than, their partner's (Brines 1993, 1994). This empirical evidence runs counter to the theoretical suggestion that work arrangements between partners are gender-neutral.

For this reason, sociologists have criticized economic theories that assume a gender-neutral housework division and have sought alternative explanations for the gender asymmetries observed in society; proposed theories include women's lack of power within the family and in society at large (Lennon and Rosenfield 1994, Baxter and Western 1998) and social gender norms, either externally imposed on women, or internalized by them (Brines 1994; Baxter and Western 1998; Bianchi et al. 2000). Hakim (2000) is the leading proponent of sociological theories based on internalized preferences. Her "preference theory" argues that preferences over

paid work and domestic work differ systematically between men (whose preferences are largely homogeneous) and women (who are highly heterogeneous). She categorizes between 10% and 30% of women as “career-oriented”, prioritizing paid work and life in the public arena; a similar proportion as “family-oriented”, prioritizing work in the home and investments in children; and the remainder as “adaptive”, valuing activity in both the domestic and the public spheres.

More recently, economists have also recognized the existence of gender norms and have attempted to identify an economic rationale for these norms. One proposal (see Hadfield 1999; Baker and Jacobsen 2007) is that a gendered division of labor attenuates co-ordination issues between partners and the “marital hold-up problem”. It is in the interests of both partners to co-ordinate the division of domestic and market labor between them, with one partner specializing more in labor market activities and the other in domestic activities. However, the co-ordination exercise may not be straightforward, especially when people have incomplete information on their partners’ characteristics. In such a situation, a customary gendered division of labor may help in attenuating these coordination issues. Additionally, men and women have to decide how much to invest in learning market and domestic activities before searching for a mate, and without information on the capabilities of their future partner in these domains. In the absence of behavioral norms, there will be an incentive for both partners to over-invest in learning market activities and to under-invest in learning domestic skills, since in the case of a divorce a partner who specializes more in market work will have better outside options than a partner who has specialized more in housework (Ott 1992). This problem of suboptimal investments is termed the marital hold-up problem², and is ameliorated by social norms relating to the gendered division of labor.

While this explanation provides a justification for these social norms, it does not explain why people conform to them. Social norms can only exist if deviating behaviors are socially sanctioned or if norm-compliance is rewarded (Axelrod 1984; Ott 1992). In modern societies there are no formal sanctions for people who deviate from a customary gender division of labor, and no formal rewards for those who conform; the only possible

² As emphasized by Baker and Jacobsen (2007), the existence of a marital hold-up problem is consistent with a large set of economic models that explain the division of marital surplus between partners and the investment decisions by partners prior to the marriage (e.g. Konrad and Lommerud 2000; Vagstad 2001; Peters and Siow 2002).

explanation for why people conform is that they internalize social norms. The internalization of social gender norms occurs when people conform to the behavior and role prescribed by social norms to affirm their gender self-image (gender identity). Behaviors and gender roles that deviate from those prescribed by social norms can cause anxiety and uneasiness and a loss of gender identity (West and Zimmerman 1987).

Akerlof and Kranton (2000) develop an economic model that takes into consideration gender identity. They assume that people choose specific actions (behaviors) to maximize their utility, which depends on the consumption of goods and services made possible by the chosen actions and on how much these chosen actions reinforce the gender identity of the individual. Individuals who choose work arrangements which deviate from arrangements prescribed by social gender norms will incur a penalty in their utility compared to individuals whose choices affirm their gender self-image by conforming to customary gender roles. This influence of gender identity on individual utility provides a potential explanation for why women do more housework than their male partners, even when their earning power and hours of market work are equal to or higher than those of their partners.

It is this idea that gender identity, in the form of internalized gender norms, influences people's preferences, and that preferences in turn influence behavior, that this paper seeks to investigate.

As mentioned in the Introduction, we are not aware of any existing research which directly examines the relationship between housework sharing within a partnership and the utility of the partners. However, several studies on related themes exist, and are relevant to this study. Gough and Killewald (2001) use a quasi-experimental approach to evaluate the causal effect of exogenous changes in the share of market work within the partnership (in the form of unexpected job losses) on housework shares. They find that the effects of job loss are not gender-neutral: both men and women increase their share of housework on losing their job, but this increase is about twice as large for women as it is for men.

Booth and Van Ours (2009) use data from the Household, Income and Labor Dynamics Survey in Australia (HILDA) to estimate the relationship between self-reported life and job satisfaction measures, and the allocation of paid work within households. They find that the men reporting the highest levels of satisfaction are those who work full-time, while the women reporting the highest levels of satisfaction are those who work part-time while their partner works full-time. However, they do not include housework shares in the model, and thus the

study does not provide any direct evidence on the relationship between gender identity and the utility derived from different housework arrangements.

Harryson et al. (2012) use a Swedish sample of cohabiting couples to study the relationship between psychological distress and different housework arrangements, finding that psychological distress is more common, in both men and women, in households where the woman does most of the housework. Sigle-Rushton (2010) finds that the incidence of divorce is lower in families where the father is involved in housework and childcare. None of these studies directly investigates the relationship between housework and individual-level utility; nevertheless, they appear to suggest that certain beneficial outcomes are associated with a more equal distribution of domestic labor.

Görges (2014) run a small lab experiment to test gender specific patterns in work specialization decisions for heterosexual couples who are partners in real life and for fictitious partners who are a pair of strangers. For each couple both individuals can choose to complete a performance-based paid task or they can cooperate and agree that one of them complete the paid task while the other performs an unpaid task that will triple the pay-rate for the partner who completes the paid task. After the tasks are completed individuals are informed about their own pay-off but not about the payoff of their partner and they can decide how much of their pay-off to share with their partner, knowing that the shared pay-off will be increased by 20% and equally distributed between the two partners. The lab experiment results seem to suggest that women are more likely to cooperate and choose the unpaid task but only if they are in a real partnership.

Several researchers (Álvarez and Miles 2003, Bitman et al. 2003; Washbrook 2007) have investigated the effect of earnings differences between partners on housework allocations. Álvarez and Miles (2003) find that gender differences are not explained by differences between partners in terms of wages or other observable characteristics, but by unobserved characteristics related to gender. Bitman et al. (2003) find that women's housework decreases as their wages increase, but only up to the point where both partners earn the same; when women earn more than men, then they appear to compensate for this deviation from gender norms by doing more of the housework. Washbrook (2007) finds that while the amount of paid work done by women, especially mothers, is related to the wage difference between partners, the labor supply of men is not. An increase in women's wages leads to a reduction in their housework and to a market

substitution of their domestic work, but this is not the case for men. All these studies suggest a degree of gender asymmetry in the relationship between wages and housework.

Hersch and Stratton (1994, 1997, 2002) and Bryan and Sevilla-Sanz (2011) also analyse the relationship between housework and wages but they focus on the effect of women's housework shares on their earnings. They provide direct evidence that the larger share of housework done by women may lead to significantly lower wages, especially if they have children.

3. DATA AND EXPERIMENTAL METHOD

3.1. The UKHLS and the Innovation Panel

This experiment was conducted as part of the UK Household Longitudinal Study (UKHLS, also known as *Understanding Society*). The UKHLS is a large-scale UK-based panel survey conducted by the Institute for Social and Economic Research at the University of Essex; it started in 2009 and has run annually since then (Buck and McFall 2012). The survey covers around 40,000 households and collects data on a range of individual and household domains; notably, for our purposes, it contains information on household structure, current and past employment, time spent on housework, individuals' standards of housework, and, for people living with a partner, the shares of housework done by respondents and their partners. Also important for this study is that fact that both members of married and cohabiting couples are eligible for interview.

A representative subset of around 1,500 households forms the survey's Innovation Panel (IP). The IP functions as a test-bed for innovations in data collection methods and new methods of research; it started in 2008, a year before the main UKHLS survey, and has been conducted each year since (Jäckle et al. 2014). IP participants are asked the same questions as other UKHLS interviewees; each year a small number of methodological experiments is also added. The experiment on housework satisfaction, on which this paper is based, forms part of the fifth IP, conducted in 2012.

3.2. Experimental design

Each individual participating in the experiment was presented with three hypothetical scenarios (vignettes) outlining different arrangements between partners for the sharing of housework.

They were then asked to indicate what their level of satisfaction would be with each of the three scenarios, on a scale from 1 (*completely dissatisfied*) to 7 (*completely satisfied*).

The three scenarios given to each respondent were selected from a battery of scenarios generated by varying five factors (“dimensions”) which are likely to impact on people’s satisfaction with housework arrangements: (1) the share of housework done by the respondent; (2) the hours of paid work of respondents and partners; (3) the hourly earnings of respondents and partners; (4) the presence and age of children in the home; and (5) whether the household employs paid help (in the form of a cleaner). Between two and five categories (“levels”) were defined for each of the five dimensions; these are presented in Table 1.

TABLE 1: DIMENSIONS AND CATEGORIES USED IN THE SCENARIOS

Dimensions	Categories				
	1	2	3	4	5
1 Hours of paid work	Resp. and partner both work full-time	Resp. and partner both work part-time	Resp. works full-time, partner works part-time	Resp. works part-time, partner works full-time	-
2 Hourly pay	Partner’s pay double that of respondent	Respondent’s pay double that of partner	Resp. and partner’s pay about equal	-	-
3 Number and age of children	No children	One child, age 6 months	One child, age 5 years	One child, age 15 years	-
4 Share of housework done by resp.	None	One quarter	Half	Three quarters	All
5 Paid housework	None	Cleaner, one morning a week	-	-	-

The full set of possible scenarios spans all 480 possible combinations³ of these categories; all experimental factors were fully crossed with each other, allowing the effects of each to be estimated free of the effects of the other categories, and also allowing estimation of the effects

³ The number of possible combinations is the product of the number of categories: $n = 4 \times 3 \times 4 \times 5 \times 2 = 480$.

of all possible interactions and trade-offs between the experimental factors. Figure 1 shows the wording of one sample scenario generated under this procedure.

FIGURE 1: SAMPLE SCENARIO, WITH THE VARIED DIMENSIONS UNDERLINED

“Imagine that you are married or cohabiting, you and your partner both have full time jobs, and your hourly pay is approximately the same as your partner’s. You have one child aged 5 years; your partner does one quarter of the housework while you do three quarters of it, and you do not employ anybody to help with the housework.”

How satisfied would you say you are with the sharing of the housework?

Completely dissatisfied	Mostly dissatisfied	Somewhat dissatisfied	Neither satisfied nor dissatisfied	Somewhat satisfied	Mostly satisfied	Completely satisfied
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The set of all 480 possible scenarios was used and split into 160 different questionnaire versions, each containing three scenarios, using a *D*-efficient sampling technique, which minimizes the correlations between dimensions (factors), and maximizes the variance of each of the factors within the questionnaire versions, therefore guaranteeing a “level balance” i.e. ensuring that each category occurs with about equal frequency (for details see Kuhfeld et al. 1994; Atzmüller and Steiner 2010; Auspurg and Hinz 2015).

These 160 questionnaire versions were randomly allocated to households participating in the experiment⁴, with the ordering of the three scenarios being randomized for each household. The randomization of question ordering neutralizes possible effects of the ordering of scenarios, such as carry-over or learning effects.

⁴ Randomization was done at the household level in order to obtain maximum statistical power when analyzing data at the partnership level. Presenting male and female partners with identical scenarios ensures that male/female differences in evaluations of the scenarios are not caused by differences in the experimental stimuli, but by differences in personal characteristics (including gender). In any case, randomizing at the household level still constitutes a random matching of experimental stimuli to personal characteristics, ensuring the high internal validity of an experimental approach (see the randomization checks below).

The experiment was administered in self-completion mode via computer-assisted self-interview (CASI)⁵. Self-completion is the recommended mode for multi-factorial experiments of this type, firstly because the scenarios may be better understood if read directly by respondents than if they are read out by an interviewer; and secondly because self-completion reduces social desirability bias (Auspurg et al. 2014).

Thorough pretests with oral feedback were run prior to the implementation of IP5, and suggested that respondents coped well with the hypothetical nature of the questions and the level of complexity of the experiment.

This experimental design has a number of advantages. The selectivity and endogeneity issues referred to earlier, which are potentially so problematic in survey-based research, do not cause problems here, since the shares of housework and paid work in the vignettes are uncorrelated with other variables in the vignettes (earnings, the presence of children, and paid help with housework) and with individuals' real-life characteristics, both observable and unobservable. The scenarios span the full space of possible combinations of housework, paid work and the other factors, so we do not run into the problem described earlier of insufficient observations in part of the space. We may be confident that the effects we estimate are indeed the effects of the allocation of housework on satisfaction, rather than a spurious effect caused by omitted variables. Because the share of housework is precisely stated in the scenarios, there are no measurement problems relating to the time spent on housework, which may be the case in surveys (see, e.g., Niemi, 1993; Lee and Waite 2005). Finally, the random allocation of stimuli to households means that all households (and all men and women in the sample) are presented on average with exactly the same scenarios. Thus, all gender differences in earning power and the shares of housework and market work have been leveled out; in this context, if we were to observe systematic gender differences in satisfaction with different types of scenario, these could not arise from different comparative advantages across the two spheres of work, as these have been effectively cancelled out in our experiment.

⁵ The main mode of data collection in the IP5 sample varied as part of the experimental design of the IP, with around two-thirds of the sample being interviewed via computer-assisted personal interview (CAPI) and the remainder completing web-based interviews. However, our experiment formed part of a self-completion module in all cases.

3.3. Respondent sample and descriptive statistics

Of 1,573 households eligible for interview at IP5, 1,224 (78% of the total) participated in the survey; 2,424 individuals in these households were eligible for personal interview, and of these, 1,995 (82% of the total) provided valid interviews. However, not all of these provided responses to the self-completion module containing the questions on housework satisfaction: in total, 1,609 of responding adults (81%) participated in the housework satisfaction experiments⁶. Some of these evaluated only one or two of the three vignettes; thus, a total of 4,547 valid evaluations were generated. Full details of sample sizes and non-response are provided in Burton (2013).

Because of non-response at various stages, it is possible that the sample of individuals providing valid responses is non-random. However, this type of experimental approach does not require a random sample of respondents, since the experimental stimuli are, by design, uncorrelated with any of the other factors affecting the dependent variable (Mutz, 2011).

Table 2 presents descriptive statistics for the sample. The mean of the dependent variable – satisfaction with the housework arrangements described in the vignettes – is just over 4, the midpoint of the range. The average satisfaction rating is slightly higher for men than for women; we will return to this in Section 4.

⁶ Note that the hypothetical nature of the vignettes, in which respondents were asked to imagine being married or living with a partner, meant that all adult IP sample members, and not just those actually living with a partner, were eligible to participate.

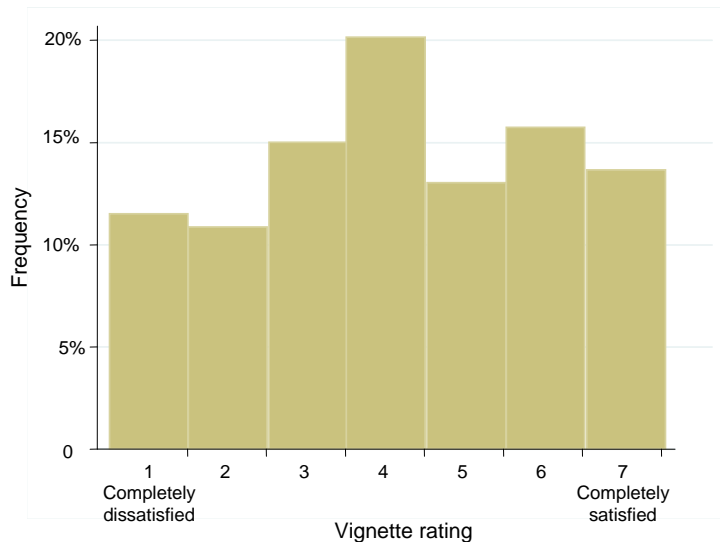
TABLE 2: DESCRIPTIVE STATISTICS

	Mean	Standard deviation
Dependent variable: self-reported satisfaction		
Women	4.042	1.919
Men	4.261	1.860
All	4.142	1.895
Vignette dimensions and categories		
Share of housework (ref: respondent does about 50%)		
Respondent does < 50% of housework	0.401	
Respondent does > 50% of housework	0.409	
Hours of paid work (ref: both partners work full-time)		
Both partners work part-time	0.248	
Respondent works full-time, partner part-time	0.250	
Respondent works part-time, partner full-time	0.241	
Hourly earnings (ref: resp. and partner earn about the same)		
Respondent earns half as much as partner	0.343	
Respondent earns twice as much	0.330	
Paid help with housework (ref: none)		
Cleaner employed one morning per week	0.506	
Presence of children (ref: none)		
6-month old child	0.236	
5-year old child	0.253	
15-year old child	0.249	
“Real-life” characteristics		
Male	0.455	
Age	47.711	16.845
Married	0.576	
Married or cohabiting	0.691	
No. of children	0.469	0.891
Working	0.599	
Student or training	0.068	
Education (ref: no qualifications)		
GCSE/O-level	0.216	
A-level	0.215	
University degree	0.280	
Other	0.189	
Number of individuals	1,609	
Number of observations	4,547	

Below this are listed the distributions of the different vignette factors. Because these have been randomized, we expect the categories in each dimension to be evenly distributed across responses, and indeed, this is the case – for example, we see that in the “*hours of paid work*” dimension, each of the four possible full-time/part-time combinations accounts for close to 25% of responses. Note that although the “*share of housework*” dimension initially consisted of five categories, we have collapsed the coding into three categories (respondent does less than 50%, about 50%, and more than 50% of the housework). This leads to much greater ease in interpreting the results later in the paper, and does not change the interpretation of the results at all.

The real-life characteristics of participants are shown in the lower part of Table 2. We see that around half the sample is male, the average age of respondents is around 48 years, and that around 70% are actually married or living with a partner.

FIGURE 2: DISTRIBUTION OF RESPONSES TO HOUSEWORK SATISFACTION VIGNETTES



Further detail on the distribution of our dependent variable is shown in Figure 2. The midpoint of the scale (neither satisfied nor dissatisfied) is the modal response; however, aside from this, the responses are much more evenly distributed across the scale, particularly at the lower end, than would be typical for “real-life” satisfaction measures carried in surveys (e.g., ONS 2013). This gives an indication that individuals are indeed responding to the wide variation in the stimuli contained in the vignettes.

3.4. Preliminary tests of validity

For the factorial survey method to work well, it is important that the three questions received by each respondent be random, in that (a) the questions received are uncorrelated with respondents’ personal characteristics; (b) the factors varying between questions are not cross-correlated; and (c) each category of each of the factors occurs with approximately equal frequency.

We checked whether these conditions held in the sample of respondents, and found that this is indeed the case. Correlation coefficients were calculated between the scenario components and eight individual- or couple-level characteristics (age, sex, marital status, number of children, actual satisfaction with housework arrangements, both partners’ hours of housework, and between-partner differences in standards of

housework). These coefficients were all below 0.04, demonstrating that condition (a) is satisfied. All cross-correlations between the factors varying between questions were also well below 0.04, satisfying condition (b). Finally, there is almost perfect balance between the levels of each of the factors, satisfying condition (c). We carry out further validity tests in Section 4.3.

4. ESTIMATION AND RESULTS

4.1. Model for estimation

Our multifactorial experimental design allows us to study the relationship between individuals' perceived utility derived from different arrangements for housework and paid work, controlling for the wage levels of both partners, for the presence and age of children, and for whether there is paid help for domestic work.

By assuming that the level of satisfaction which people report for different hypothetical scenarios reflects their actual utility, say y^* , we estimate the following utility model:

$$y_{is}^* = X_{is}\beta + \mu_i + \varepsilon_{is}, \quad (1)$$

where y_{is}^* is the utility of individual i corresponding to the vignette (scenario) s ($s=1,2,3$); X_{is} is the vector of explanatory variables that characterize the vignette's factors; μ_i is the individual-specific effect capturing characteristics that are specific to the individual and which might affect the level of reported satisfaction (for example, personality traits and mood on the day of the interview); and ε_{is} is the idiosyncratic error term which we assume to be independent of the explanatory variables.

We begin by estimating a linear regression model with random effects, with robust standard errors to take account of the correlation in the error term within individuals across vignettes. The model includes dummy variables for different levels of the five factors describing the vignettes, plus interactions between the housework share dummies and each of the remaining four factors; we also control for individual age and age squared.

4.2. Main results

Results from linear random effects models are presented in Table 3. Main effects from the vignette factors are reported at the top of the table, followed by interactions between housework shares and all the other factors. Because this is a very large table, only coefficients and significance levels are shown; full results with standard errors are reported in Table A1 in Appendix A.

TABLE 3: LINEAR RANDOM EFFECTS MODEL OF SATISFACTION RATINGS

	(1)	(2)	(3)	(4)
	All	Women	Men	Gender difference
Main effects: vignette factors				
Share of housework (ref: about the same)				
Resp. does < 50% of housework	-1.106***	-1.194***	-1.012***	-0.190
Resp. does > 50% of housework	-1.539***	-1.699***	-1.387***	-0.306
Hours of paid work (ref: both full-time)				
Both part-time	-0.167	-0.188	-0.094	-0.096
Resp. full-time, partner part-time	-0.726***	-1.030***	-0.365*	-0.671**
Resp. part-time, partner full-time	-0.287*	-0.355	-0.179	-0.166
Hourly earnings (ref: about the same)				
Resp. hourly wage half that of partner	-0.033	0.015	-0.060	0.063
Resp. hourly wage double partner's	-0.148	-0.101	-0.158	0.059
Paid help with housework (ref: none)				
Cleaner one morning per week	-0.014	-0.015	-0.050	0.027
Presence of children (ref: none)				
6-month-old child	-0.005	0.157	-0.238	0.394
5-year-old child	0.034	0.306	-0.273	0.585**
15-year-old child	0.322**	0.315	0.307	0.001
Interactions: Resp. does <50% housework X				
Both part-time	0.313	0.337	0.238	0.106
Resp. full-time, partner part-time	1.249***	1.537***	0.921***	0.633*
Resp. part-time, partner full-time	0.533***	0.460*	0.571**	-0.113
Resp. hourly wage half that of partner	0.009	-0.082	0.077	-0.154
Resp. hourly wage double partner's	0.438***	0.458**	0.352	0.091
Cleaner one morning per week	0.105	0.092	0.153	-0.062
6-month-old child	-0.034	-0.057	0.066	-0.097
5-year-old child	-0.118	-0.333	0.157	-0.494
15-year-old child	-0.578***	-0.519**	-0.601**	0.100
Interactions: Resp. does >50% housework X				
Both part-time	0.088	0.018	0.156	-0.131
Resp. full-time, partner part-time	0.531***	0.771***	0.261	0.505
Resp. part-time, partner full-time	1.062***	1.313***	0.760***	0.529
Resp. hourly wage half that of partner	0.167	0.049	0.299	-0.244
Resp. hourly wage double partner's	0.084	-0.030	0.180	-0.217
Cleaner one morning per week	0.352***	0.453**	0.256	0.198
6-month-old child	-0.034	-0.178	0.218	-0.397
5-year-old child	-0.079	-0.370	0.282	-0.650*
15-year-old child	-0.391**	-0.358	-0.381	0.038
Age	-0.044***	-0.043***	-0.040**	-0.003
Age squared	0.000***	0.000***	0.000***	-0.000
Constant	5.875***	5.933***	5.733***	0.205
Number of observations	4,547	2,476	2,071	4,547
Number of individuals	1,541	841	700	1,541
Wald test χ^2	425.59	312.33	158.11	479.93
p-value	0.000	0.000	0.000	0.000
Rho	0.473	0.434	0.518	0.472

Notes: Asterisks denote significance levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(1) Model with equal coefficients across gender and using the pooled sample of women and men,

(2) Separate model for women, (3) separate model for men,

(4) Model with different coefficients across gender and using the pooled sample of women and men.

The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(31)$ for Columns 1-3 and $\chi^2(63)$ for Column 4.

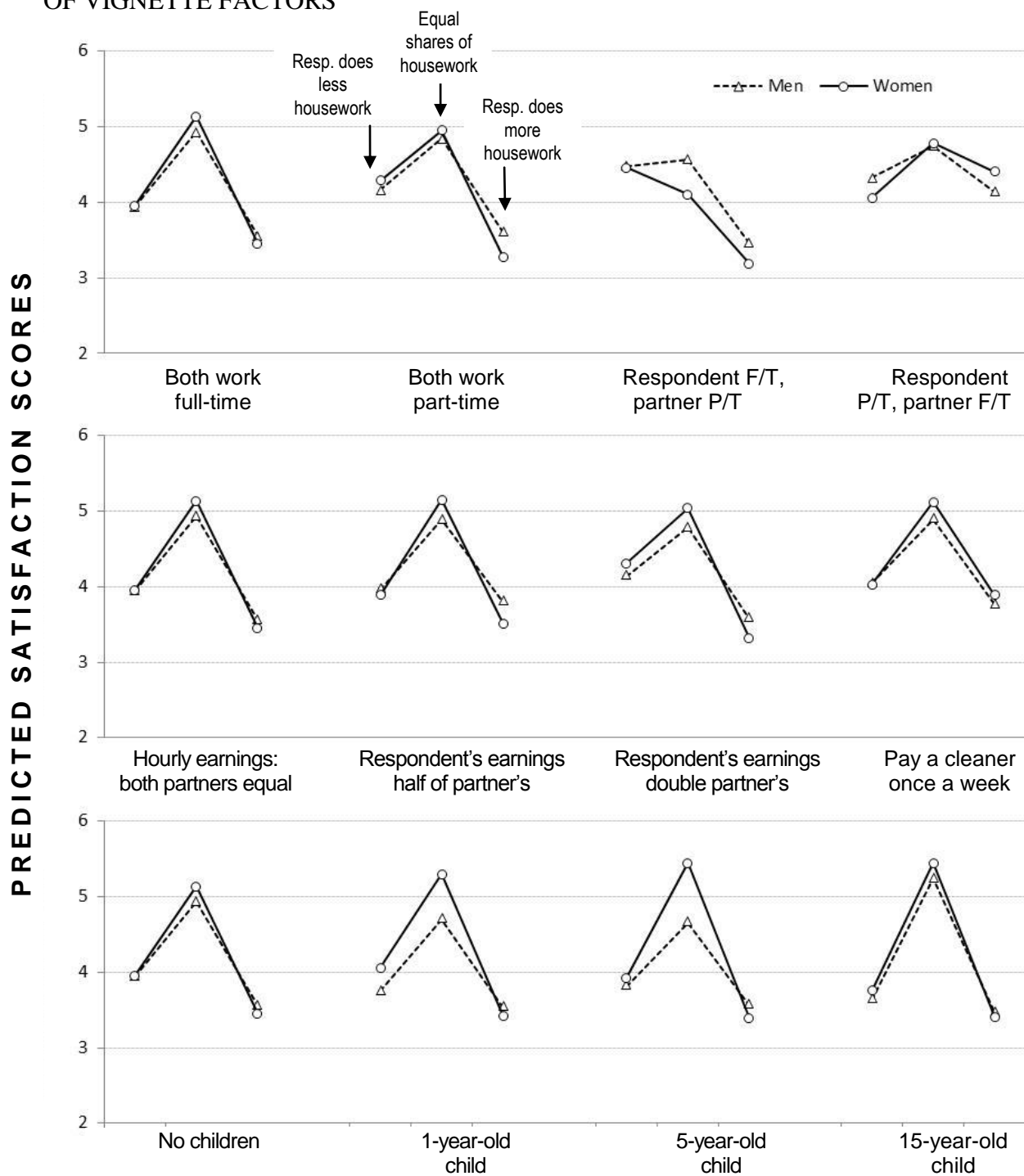
Rho is the fraction of the variance of the unobserved component explained by the random effect.

Results are shown for the combined sample of men and women (Column 1) and for women and men separately (Columns 2 and 3); Column 4 shows the difference in the coefficients across gender computed using the pooled sample of women and men but allowing the coefficients to vary by gender⁷. The main effects suggest that both men and women have a preference for housework to be distributed equally between the members of a couple, with both the alternatives (doing less housework than one's partner, and doing more than one's partner) being associated with significantly lower satisfaction scores. The negative coefficients on unequal hours of housework are offset (but only partially) by the interaction terms between housework and shares of paid work. These interactions provide additional insights into the general preference for equity. In the first group of interactions (interactions with the "respondent does < 50% housework" variable), the interaction coefficient on "respondent works full-time, partner works part-time" is much larger than the corresponding coefficient on "respondent part-time, partner full-time"; that is, people are happier doing less housework if they do more paid work than their partner. In the second group (interactions with the "respondent does > 50% housework" variable, the interaction coefficient on "respondent part-time, partner full-time" is much larger than the interaction coefficient on "respondent full-time, partner part-time"; that is, people are happier doing more housework if they do less paid work than their partner. Thus, as well as indicating a preference for an equal division of housework, the results also suggest that respondents are considering the total distribution of paid and unpaid work, and indicating a preference for equity in this total allocation.

These coefficients are most easily interpreted graphically. Figure 3 shows predicted satisfaction levels for men and women, varying the vignette factors.

⁷ The differences in coefficients reported in Column 4 are almost identical to the differences between Columns 2 and 3; the slight discrepancies are due to small differences in the variance of the unobserved component between men and women.

FIGURE 3: PREDICTED SATISFACTION SCORES FOR DIFFERENT COMBINATIONS OF VIGNETTE FACTORS



Notes: Each set of predicted values contains three pairs of predictions:
 Respondent does LESS housework than partner (left-hand side)
 Respondent and partner do EQUAL SHARES of housework (center)
 Respondent does MORE housework than partner (right-hand side).

Predicted satisfaction scores are estimated from pooled sample of men and women, imposing zero random effect and zero idiosyncratic error. Predictions based on individuals aged 48, with characteristics in reference category unless otherwise specified.

Predicted values are calculated at the midpoint of the age variable (48 years) and, unless stated, for the reference groups for all the vignette variables. Predictions are estimated using the pooled sample of men and women, imposing a zero random effect and a zero idiosyncratic error ε_{is} .

Three pairs of predicted values are shown for each situation. In each case, the left-hand pair relates to the situation in which the respondent does less housework than his or her partner; the central pair relates to the situation in which both partners do equal shares, and the right-hand pair relates to the situation in which the respondent does more housework.

In general, it is evident that both men and women prefer a situation in which the housework is equally shared. In most scenarios, both men and women seem somewhat to prefer to have their partner do most of the housework over doing most of the housework themselves; however, these differences are small compared to the substantial differences in preferences between equal and unequal shares.

The only situation in which there is not an unequivocal preference for equal shares of housework is when the respondent works full-time and their partner works part-time. In this case, both men and women (not unreasonably) are considerably less happy with a situation in which they do most of the housework, than in a situation in which they take a minor share.

In virtually all scenarios, it is clear that the differences between predicted values for men and women are very small. For the sake of clarity, confidence intervals have not been shown on the graph (they are available on request), but most of the gender differences fall well short of statistical significance, even at the 10% level. There are two exceptions: when a 1-year-old or 5-year-old child is present, women seem to be happier with equal shares of the housework than when no child is present, while men seem to be less happy. While the point estimates for these gender differences are significant at the 10% level, they do not change our over-arching finding, namely that the structures of men's and women's preferences over the distribution of housework are extremely similar. It is likely to be these two situations which are the main drivers for the slightly lower mean satisfaction of women reported in Table 2; Column 4 of Table 3 provides no evidence that the constant terms differ significantly by gender.

Testing formally for gender differences across the full model allowing all coefficients to differ between men and women, a Wald test (distributed as χ^2 with 32 degrees of freedom) takes the value 45.05, with a p -value of 0.063. Thus, at the 5% level of significance, we do not reject the hypothesis that all regression coefficients between genders are equal.

These results suggest that the structure of preferences does not differ systematically between men and women; in fact, with one or two exceptions, men and women have remarkably similar preferences over the allocation of paid work and housework.

In particular, there is no evidence at all that women have stronger preferences than men for a larger share of housework or for a smaller share of market work (or, put another way, men do not have stronger preferences than women for a smaller share of the housework or for a larger share of market work). The main finding is of a preference for equity: both sexes appear to prefer an equal allocation of housework, and both sexes are more likely to feel more favorably disposed to doing a higher share of the housework if their partners are doing more of the market work. This suggests that the gendered division of labor which we observe in real life cannot be explained by gender differences in preferences or by internalization of gender norms, but must be caused by some other mechanism.

4.3. Validity and robustness checks

The linear model estimated above implicitly assumes that reported satisfaction, which is rated on a 7-point scale, is a direct measure of utility, and that each increment on the 7-point scale corresponds to a similar increment in individuals' utility. If this assumption does not hold, the linear model may give biased results. As a check on this, we re-estimate the model using an ordered probit specification; estimated coefficients are reported in Table 4. Results are very similar to those reported in Table 3; in particular, the estimated coefficients are extremely similar between men and women (standard errors are reported in Table A2 in Appendix A).

TABLE 4: ORDERED PROBIT MODEL FOR SATISFACTION WITH INDIVIDUAL
RANDOM EFFECTS

Variables	(1) All	(2) Women	(3) Men	(4) Gender difference
Main effects: vignette factors				
Share of housework (ref: about the same)				
Resp. does < 50% of housework	-0.925***	-0.973***	-0.898***	-0.169
Resp. does > 50% of housework	-1.270***	-1.355***	-1.217***	-0.258
Hours of paid work (ref: both full-time)				
Both part-time	-0.149	-0.178	-0.080	-0.106
Resp. full-time, partner part-time	-0.621***	-0.843***	-0.348	-0.554*
Resp. part-time, partner full-time	-0.266*	-0.308	-0.183	-0.135
Hourly earnings (ref: about the same)				
Resp. hourly wage half that of partner	-0.035	-0.003	-0.057	0.041
Resp. hourly wage double partner's	-0.136	-0.090	-0.151	0.054
Paid help with housework (ref: none)				
Cleaner one morning per week	-0.005	-0.040	-0.005	-0.047
Presence of children (ref: none)				
6-month-old child	-0.024	0.102	-0.233	0.325
5-year-old child	0.013	0.247	-0.269	0.517*
15-year-old child	0.276*	0.314	0.217	0.117
Interactions: Resp. does <50% housework X				
Both part-time	0.262	0.300	0.190	0.131
Resp. full-time, partner part-time	1.021***	1.210***	0.808***	0.520
Resp. part-time, partner full-time	0.456**	0.383	0.507*	-0.085
Resp. hourly wage half that of partner	0.019	-0.043	0.070	-0.110
Resp. hourly wage double partner's	0.367**	0.366*	0.313	0.070
Cleaner one morning per week	0.068	0.084	0.089	0.004
6-month-old child	0.011	-0.015	0.120	-0.104
5-year-old child	-0.082	-0.279	0.184	-0.468
15-year-old child	-0.478**	-0.461*	-0.470*	-0.021
Interactions: Resp. does >50% housework X				
Both part-time	0.094	0.040	0.158	-0.112
Resp. full-time, partner part-time	0.457**	0.638***	0.246	0.420
Resp. part-time, partner full-time	0.860***	1.013***	0.658**	0.406
Resp. hourly wage half that of partner	0.135	0.054	0.243	-0.166
Resp. hourly wage double partner's	0.094	0.013	0.166	-0.153
Cleaner one morning per week	0.268**	0.363**	0.181	0.216
6-month-old child	-0.002	-0.116	0.229	-0.338
5-year-old child	-0.061	-0.310	0.263	-0.565*
15-year-old child	-0.346*	-0.382	-0.275	-0.125
Age	-0.037***	-0.035**	-0.0359*	-0.002
Age squared	0.000***	0.000**	0.000**	0.000
Observations	4,547	2,476	2,071	4,547
Number of individuals	1,541	841	700	1,541
Wald test χ^2	1006.78	440.52	570.49	1000.92
Wald p -value	0.000	0.000	0.000	0.000
Rho	0.547	0.496	0.605	0.548

Notes: Asterisks denote significance levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(1) Model with equal coefficients across gender and using the pooled sample of women and men,

(2) Separate model for women, (3) separate model for men,

(4) Model with different coefficients across gender and using the pooled sample of women and men.

The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(31)$ for Columns 1-3 and $\chi^2(63)$ for Column 4.

Rho is the fraction of the variance of the unobserved component explained by the random effect.

The principal advantage of our factorial experiment is that it provides a solution to the problems of sample selection and endogeneity which may give rise to bias in satisfaction models when data from standard sample surveys is used. This is particularly important when attempting to identify the role of gender as a determinant of satisfaction with working arrangements between partners, since in real life these arrangements are not randomly allocated between men and women. However, a number of potential threats to the validity of our experiment remain. First, the fact that vignettes (and factors within vignettes) have been randomly assigned to sample members, does not necessarily mean that their distribution is completely random among respondents, since some people do not provide a response to one or more of the vignette questions.

A second possible threat to the validity of our experiment is that our estimates may be biased if self-reported measures of satisfaction are not comparable between individuals because they are influenced by personal perceptions and emotional factors such as a respondent's state of mind at the time of interview.

Finally, the estimated effects of gender on satisfaction with housework may be inaccurate if people who are asked to report their level of satisfaction with different hypothetical work arrangements give answers that abstract from their own gender.

We address the first potential problem by checking whether our final subsample still constitutes a valid random experiment; to do this, we need to confirm that the factors defining each vignette are uncorrelated with both observed and unobserved individual characteristics. It is relatively straightforward to test for correlation with observed characteristics (see the validity tests described in Section 3.2). We test for correlation with unobserved characteristics by estimating model (1) under both fixed and random effects specifications, and testing whether there are statistically significant differences between the two sets of estimates, using a Hausman-type test that is robust to potential correlation in the errors within individuals, as suggested by Wooldridge (2002); in effect, this constitutes a test of whether the controls in the model are correlated with individual-level unobservable characteristics. The computed p -value of this test is 0.401; we therefore do not reject the hypothesis that the coefficients from fixed and random effects models are equal, and we do not reject the hypothesis that the factors are indeed randomly assigned among respondents. We repeat this test with an extended version of model (1), including as additional explanatory variables individuals' real-life level of education, job status, marital status and number of children. The results are similar, confirming that the

vignette factors are indeed uncorrelated with both observable and unobservable individual characteristics, and that our experiment is a valid random experiment.

The second potential problem is also addressed by the comparison of fixed and random effects coefficients from model (1). Individual-level factors such as personality traits or an individual's mood at the time of completing the questionnaire may affect the ways in which satisfaction is reported: that is, they may lead to individuals who experience similar levels of satisfaction with a particular arrangement to report their levels of satisfaction differently. Fixed effects estimates net out these between-individual differences; the fact that there are no statistically significant differences between our fixed and random effects estimates demonstrates that to the extent that there exist unobserved factors which cause individuals to report their levels of satisfaction differently, these cannot bias our results because they are uncorrelated with our vignette factors. In real life, of course, factors describing partners' work arrangements, earning power and so on are not generally uncorrelated with other individual characteristics, and this *can* lead to biased estimates.

Finally, to check the third issue – that is, whether the absence of gender differences is caused by people abstracting from their own gender when reporting their levels of satisfaction for different vignettes – we carry out a sensitivity analysis using a subsample of people for whom gender identity had been “primed” via a set of questions administered prior to the vignettes, which asked individuals to report their satisfaction with a different life domain by comparing themselves with other people of their own gender (see Burton 2013 for details). When we estimate model (1) with individual random effects for this subsample, and test between-gender differences, we do not reject the equality of the coefficients between genders (p -value 0.786). This suggests that the lack of significant gender differences in satisfaction with housework is genuine rather than being caused by people abstracting from their own gender when people answer the vignette questions.

In our experiment we assume that people abstract from their actual work arrangements and from the presence or absence of a partner or of children in their real lives when rating hypothetical scenarios. But this assumption does not automatically hold, and events such as partnership formation or the birth of children, or real-life housework arrangements, may influence the relationship between gender and preferences. Thus, although we observe very few differences between men's and women's preferences across the whole sample, it is possible that there may be gender differences in certain subgroups of the population. To test

this, we repeat our analysis on the subgroups of people who (a) actually do, or do not, live with a spouse or partner; (b) actually do, and do not, live with children; and (c) report actually doing at most half, or less than half, the housework. For each of these subgroups we re-estimate the linear regression model in Table (3) allowing the coefficients to differ between men and women, and test for the equality of these coefficients across gender using a Wald test, distributed as chi-squared with 32 degrees of freedom.

There appear to be no differences between the sexes by partnership status: the p -value for the equality of coefficients across genders under the Wald test is 0.227 for people who are married or living with a partner, and 0.133 for single people.

The analogous test, across people who do and do not have children, yields p -values for the equality of coefficients across genders of 0.454 for people without children and 0.054 for people with children. This gender difference among people with children, significant at the 10% but not the 5% level, arises partly from the fact that in this group, women show more of an aversion to unequal housework arrangements (in both directions) than men do; there are also differences between men and women in the coefficient on doing more paid work than one's partner (with women showing a greater aversion to this arrangement), although these cancel out when interaction terms are considered.

Differences between the samples of people doing at most half, and more than half, the housework follow a similar pattern: among people who do at most half of the housework in real life, there are no significant gender differences (p -value 0.390), whereas among people who actually do more than half the housework, the gender differences are significant at the 10% level (p -value 0.061). In this case, the differences are driven mainly by differences in the effects of a five-year-old child (which leads to higher satisfaction of women compared to men), which cancel out when interaction terms are considered; there do not appear to be systematic differences between men and women over preferences for housework or paid work.

Thus, at the 5% level of significance we do not reject the hypothesis that men's and women's preferences for work arrangements are identical in any of these subsamples, while at the 10% level of significance, there is some evidence of gender differences among people who actually have children, and among people who do more than half the housework in real life, although these do not appear to relate primarily to preferences over housework.

One final test relates to the practice of “satisficing” in surveys (Krosnick et al. 1996; Oppenheimer et al. 2009). Satisficing is the practice whereby some respondents provide answers to questions, but shortcut their cognitive efforts, to the extent that their responses do not correctly reflect their situation or their opinions. One common strategy of “satisficers” is to give identical answers to all items in a battery of questions; these are often located at the midpoint of the range. We repeated our analysis excluding respondents who had given the same response to all three vignettes (note that not all of these will have been “satisficers”, since some people will be genuinely indifferent between the vignettes). Our results on gender differences were not changed.

5. CONCLUSIONS

In this paper, we have discussed competing theories which seek to explain the highly gendered distribution of housework within couples. We have focused on one group of theories, namely those which propose that gender identity (that is, internalized gender norms) is responsible for gendered housework shares, via its effect on men’s and women’s relative preferences for housework and paid work. We test this theory in an experimental context, by estimating the levels of satisfaction with different allocations of paid and unpaid work within a partnership, and the differences between men and women in these levels of satisfaction.

We ask people to visualize themselves in a range of hypothetical scenarios in which the share of housework and paid work done by the respondent and his or her partner is varied, and to report on how satisfactory they find each of these scenarios. Our experimental design circumvents a range of problems which may occur if one attempts to estimate the utility which men and women derive from different work allocations using survey data: post-hoc rationalization, endogeneity, and a paucity of observations on non-standard work arrangements.

Our main finding is that both men and women display a marked preference for equity within a partnership, in terms of both the allocation of housework and the total allocation of paid and unpaid work; there is little evidence that either men or women are systematically selfish in their preferences, or that men’s preferences differ systematically from those of women, or that either men or women prefer arrangements under which the woman specializes in home production while the man specializes in market work. We are not aware of any existing studies with which our findings may be directly compared; however, our results are

certainly congruent with those of Harryson et al. (2012), which suggest that levels of psychological distress are lower, among both men and women, in partnerships in which housework is shared equally; and those of Sigle-Rushton (2010) who finds that the incidence of divorce is lower among couples who share housework and childcare.

It is clear from previous studies (Álvarez and Miles, 2003; Washbrook, 2007, and others) that the higher share of housework done by women can only be explained partially by gender differences in observable characteristics. Our findings add to this by indicating that it is also unlikely that gender differences in housework shares can be explained by systematic differences in the utility that men and women derive from doing housework.

So, why *do* women do such a large share of the housework? It is worth noting that our finding that there are no systematic gender differences in preferences over housework allocations does *not* mean that there are no systematic gender differences in preferences over other domains, or in personality traits. For example, several psychological studies (Costa et al. 2001; Schmitt et al. 2008) have reported that women score more highly than men on the personality dimension of agreeableness. In the presence of a marital hold-up problem, women's tendency to be more agreeable and less antagonistic (see e.g. Bertrand 2011) may mean they end up investing more in housework, even if this is not economically the best choice for them. Put another way, even though women do not derive any more utility from doing housework than men do, they may derive a greater level of utility than men from avoiding conflict in a relationship, with the net result that they end up doing more housework.

In this article we have highlighted the multiple advantages of a vignette-based experimental approach for improving our understanding of the determinants of the gendered distribution of housework; there is no reason why similar experimental techniques could not be used to examine the possible role of unobservable differences in personality and preferences over other domains.

References

- Abraham M., K. Auspurg, and T. Hinz (2010). Migration Decisions Within Dual-Earner Partnerships: A Test of Bargaining Theory. *Journal of Marriage and Family* 72 (4): 876-892.
- Akerlof G. A., and R. E. Kranton (2010). Economics and Identity. *Quarterly Journal of Economics* 115 (3): 715-753.
- Álvarez B., and D. Miles (2003). Gender Effect on Housework Allocation: Evidence from Spanish Two-Earner Couples. *Journal of Population Economics* 16(2): 227-242.
- Amaya-Amaya M., K. Gerard, and M. Ryan (2008). Discrete Choice Experiments in a Nutshell, pp. 13-46 in: Ryan, M., K. Gerard, and M. Amaya-Amaya (eds.), *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht, NL: Springer.
- Atzmüller, C., and P. M. Steiner (2010). Experimental Vignette Studies in Survey Research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6(3): 128-138.
- Auspurg K., and T. Hinz (2015). *Factorial Survey Experiments. Series Quantitative Applications in the Social Sciences No. 175*. Thousand Oaks, CA: Sage.
- Auspurg, K., T. Hinz, C. Sauer, and S. Liebzig (2014). The Factorial Survey as a Method for Measuring Sensitive Issues, pp. 137–149 in: U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis (eds.), *Improving Survey Methods: Lessons From Recent Research*. New York: Routledge.
- Axelrod R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Baker M. J., and J.P. Jacobsen (2007). Marriage, Specialization, and the Gender Division of Labor. *Journal of Labor Economics* 25 (4) 763-793.
- Bateman, I. J., R. T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Ozdemiroglu, D. Pearce, J. Sugden, and J. Swanson (2002). *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham, UK: Edward Elgar.
- Baxter, J., and M. Western (1998). Satisfaction with Housework: Examining the Paradox. *Sociology* 32(1): 101-120.
- Becker G. S. (1965). A Theory of the Allocation of Time. *Economic Journal*, 75, 493–517.

- Becker G. S. (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics* 3 (1): S3-S58.
- Bertrand M. (2011). New Perspectives on Gender, pp. 1545-1592 in: O. Ashenfelter, and D. Card (eds.), *Handbook of Labor Economics*. edition 1, vol. 4b. Amersdam: Elsevier.
- Bertrand M., and S. Mullainathan (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *The American Economic Review* 91(2): 67-72.
- Bianchi S. M., M. A. Milkie, L. C. Sayer, and J. P. Robinson (2000). Is Anyone Doing the Housework? Trends in the Gender Division of Household Labor. *Social Forces* 79(1): 191-228.
- Bitman M., P. England, L. Sayer, N. Folbre, and G. Matheson (2003). When Does Gender Trump Money? Bargaining and Time in Household Work. *American Journal of Sociology* 109 (1): 186–214.
- Booth A. L., and J. C. Van Ours (2009). Hours of Work and Gender Identity: Does Part-Time Work Make the Family Happier? *Economica* 76 (301): 176-196.
- Brines, J. (1993). The Exchange Value of Housework. *Rationality and Society* 5 (3): 302-340.
- Brines, J. (1994). Economic Dependency, Gender, and the Division of Labor at Home. *American Journal of Sociology* 100 (3): 652-688.
- Bryan M. L., and A. Sevilla-Sanz (2011). Does Housework Lower Wages? Evidence for Britain. *Quarterly Journal of Economics* 63 (1): 187-210.
- Buck, N., and S. McFall (2012). Understanding Society: Design Overview. *Longitudinal and Life Course Studies* 3(1): 5-17.
- Burton, J. (2013) (ed.). Understanding Society Innovation Panel Wave 5: Results from Methodological Experiments. *Understanding Society Working Paper Series* 2013-06. University of Essex: Institute for Social and Economic Research.
- Cadsby C.B., Servatka M., and F. Song (2013). How Competitive are Female Professionals? A Tale of Identity and Conflict. *Journal of Economic Behavior & Organization* 92 (C): 284-303.

- Costa Jr P., A. Terracciano, and R. R. McCrae (2001). Gender Differences in Personality Traits Across Cultures: Robust and Surprising Findings. *Journal of Personality and Social Psychology* 81(2): 322-331.
- Croson R., and U. Gneezy (2009). Gender Differences in Preferences. *Journal of Economic Literature* 47 (2): 448-474.
- Görge L. (2014). The Power of Love: A subtle Driving Force for Unequalitarian Labor Division? *WiSo-HH Working Paper Series*, Working Paper No. 17, forthcoming, *the Review of the Economics of the Household*.
- Gough M., and A. Killewald (2011). Unemployment in Families: The Case of Housework. *Journal of Marriage and Family* 73 (5): 1085-1100.
- Green, P. E., and V. Srinivasan (1990). Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice. *Journal of Marketing* 54 (4): 3-19.
- Hadfield, G.K. (1999). A Coordination Model of the Sexual Division of Labor. *Journal of Economic Behavior and Organization*, 40 (2): 125-153.
- Hakim, C. (1996). The Sexual Division of Labour and Women's Heterogeneity. *British Journal of Sociology* 47(1): 178-188.
- Hakim C. (2000). *Work-Lifestyle Choices in the 21st Century: Preference Theory*. Oxford: Oxford University Press.
- Harryson, L., M. Novo, and A. Hammarström (2012). Is Gender Inequality in the Domestic Sphere Associated with Psychological Distress Among Women and Men? Results from the Northern Swedish Cohort. *Journal of Epidemiology and Community Health* 66(3): 271-276.
- Hersch, J. and L. S. Stratton (1994). Housework, Wages, and the Division of Housework Time for Employed Spouses. *The American Economic Review, Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association* 82(2): 120-125.
- Hersch, J., and L. S. Stratton (1997). Housework, Fixed Effects, and Wages of Married Workers. *Journal of Human Resources* 32(2): 285-307.
- Hersch, J. and L. S. Stratton (2002), Housework and Wages. *The Journal of Human Resources* 37 (1): 217-229.

- Jäckle, A., J. Burton, O. Kaminska, P. Lynn, S. McFall, and N. Uhrig (2014). *Understanding Society: The UK Household Longitudinal Study Innovation Panel, Waves 1-6, User Manual*. University of Essex: Institute for Social and Economic Research.
- Kan, M. Y., O. Sullivan, and J. Gershuny (2011). Gender Convergence in Domestic Work: Discerning the Effects of Interactional and Institutional Barriers From Large-Scale Data. *Sociology* 45(2): 234-251.
- Konrad, K. A., and K. E. Lommerud (2000). The bargaining family revisited. *Canadian Journal of Economics*, 33(2):471–87.
- Krosnick, J. A., Narayan, S., and W. R. Smith (1996). Satisficing in Surveys: Initial Evidence. *New Directions for Evaluation* 1996(70): 29-44.
- Kuhfeld, W. F., R. D. Tobias, and M. Garrat (1994). Efficient Experimental Design With Marketing Research Applications. *Journal of Marketing Research* 31 (4): 545-557.
- Lachance-Grzela, M., and G. Bouchard (2010). Why Do Women Do the Lion's Share of Housework? A Decade of Research. *Sex Roles* 63(11-12): 767-780.
- Lee, Y.-S., and L. J. Waite (2005). Husbands' and Wives' Time Spent on Housework: A Comparison of Measures. *Journal of Marriage and Family* 67(2): 328-336.
- Lennon, M. C., and S. Rosenfield (1994). Relative Fairness and the Division of Housework: The Importance of Options. *American Journal of Sociology* 100(2): 506-531.
- Louviere, J. J., D. A. Hensher, and J. D. Swait (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge Press.
- Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press.
- Niemi, I. (1993). Systematic Error in Behavioral Measurement: Comparing Results From Interview and Time Budget Studies. *Social Indicators Research* 30 (2/3): 229-244.
- Oakley, A. (1974). *The Sociology of Housework*. New York: Pantheon Press.
- ONS (2013). *Personal Wellbeing Across the UK, 2012-13*. UK: Office for National Statistics.
- Oppenheimer, D. M., T. Meyvis, and N. Davidenko (2009). Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power. *Journal of Experimental Social Psychology* 45(4): 867-872.

- Ott N. (1992). *Intrafamily Bargaining and Household Decisions*. Berlin: Springer.
- Peters, M., and A. Siow (2002). Competing premarital investments. *Journal of Political Economy*, 110(3): 592–608.
- Riach P. A., and J. Rich (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal* 112: F480-F518.
- Sándor Z., and P. H. Franses (2009). Consumer Price Evaluations Through Choice Experiments. *Journal of Applied Econometrics* 24 (3): 517–535.
- Schmitt, D. P., A. Realo, M. Voracek, and J. Allik (2008). Why Can't a Man be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *Journal of Personality and Social Psychology* 94(1): 168-182.
- Sigle-Rushton, W. (2010). Men's Unpaid Work and Divorce: Reassessing Specialization and Trade in British Families. *Feminist Economics* 16(2): 1-26.
- Stratton, L. S. (2012). The Role of Preferences and Opportunity Costs in Determining the Time Allocated to Housework. *The American Economic Review* 102(3): 606-611.
- Vagstad, S. (2001). On private incentives to acquire household production skills. *Journal of Population Economics*, 14(2): 301–12.
- Wallander L. (2009). 25 Years of Factorial Surveys in Sociology: A Review. *Social Science Research* 38: 505–520.
- Washbrook E. (2007). Explaining the Gender Division of Labour: The Role of the Gender Wage Gap. *Centre for Market and Public Organisation, CMPO Working Paper No. 07/174*.
- West C., and D.H. Zimmerman (1987). Doing Gender. *Gender & Society* 1(2): 125-151.
- Wooldridge J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

APPENDIX: TABLES OF MAIN RESULTS WITH STANDARD ERRORS

TABLE A1 LINEAR RANDOM EFFECTS MODEL OF SATISFACTION RATINGS WITH STANDARD ERRORS

VARIABLE	(1) All	(2) Women	(3) Men	(4) Gender difference
Lower housework share	-1.106*** (0.203)	-1.194*** (0.301)	-1.012*** (0.272)	-0.190 (0.406)
Higher housework share	-1.539*** (0.193)	-1.699*** (0.282)	-1.387*** (0.265)	-0.306 (0.387)
Both partners work part-time (Both PT)	-0.167 (0.163)	-0.188 (0.243)	-0.094 (0.215)	-0.096 (0.325)
Respondent work full-time, partner part-time (FT-PT)	-0.726*** (0.159)	-1.030*** (0.231)	-0.365* (0.205)	-0.671** (0.310)
Respondent work part-time, partner full-time (PT-FT)	-0.287* (0.166)	-0.355 (0.240)	-0.179 (0.228)	-0.166 (0.331)
Respondent earns half (Earns half)	-0.033 (0.128)	0.015 (0.187)	-0.060 (0.171)	0.063 (0.255)
Resp. earns twice as much (Earns twice)	-0.148 (0.130)	-0.101 (0.184)	-0.158 (0.181)	0.059 (0.259)
Pay for housework	-0.014 (0.106)	-0.015 (0.158)	-0.050 (0.137)	0.027 (0.209)
6-month old child	-0.005 (0.154)	0.157 (0.227)	-0.238 (0.199)	0.394 (0.303)
5-year old child	0.034 (0.147)	0.306 (0.218)	-0.273 (0.195)	0.585** (0.293)
15-year old child	0.322** (0.154)	0.315 (0.225)	0.307 (0.206)	0.001 (0.305)
Lower housework share Both PT	0.313 (0.192)	0.337 (0.278)	0.238 (0.264)	0.106 (0.383)
Lower housework share FT-PT	1.249*** (0.190)	1.537*** (0.267)	0.921*** (0.264)	0.633* (0.376)
Lower housework share PT-FT	0.533*** (0.194)	0.460* (0.273)	0.571** (0.274)	-0.113 (0.387)
Lower housework share Earns half	0.009 (0.155)	-0.082 (0.228)	0.077 (0.202)	-0.154 (0.307)
Lower housework share Earns twice	0.438*** (0.157)	0.458** (0.222)	0.352 (0.220)	0.091 (0.313)
Lower housework share Pay for housework	0.105 (0.128)	0.092 (0.186)	0.153 (0.175)	-0.062 (0.255)
Lower housework share 6-month old child	-0.034 (0.185)	-0.057 (0.267)	0.066 (0.253)	-0.097 (0.369)
Lower housework share 5-year old child	-0.118 (0.177)	-0.333 (0.261)	0.157 (0.233)	-0.494 (0.350)
Lower housework share 15-year old child	-0.578*** (0.177)	-0.519** (0.250)	-0.601** (0.250)	0.100 (0.353)
Larger housework share Both PT	0.088 (0.191)	0.018 (0.282)	0.156 (0.251)	-0.131 (0.378)
Larger housework share FT-PT	0.531*** (0.187)	0.771*** (0.272)	0.261 (0.245)	0.505 (0.367)

Larger housework share	1.062***	1.313***	0.760***	0.529
PT-FT	(0.197)	(0.284)	(0.270)	(0.392)
Larger housework share	0.167	0.049	0.299	-0.244
Earns half	(0.156)	(0.222)	(0.217)	(0.311)
Larger housework share	0.084	-0.030	0.180	-0.217
Earns twice	(0.161)	(0.223)	(0.230)	(0.320)
Larger housework share	0.352***	0.453**	0.256	0.198
Paid housework	(0.125)	(0.184)	(0.165)	(0.247)
Larger housework share	-0.034	-0.178	0.218	-0.397
6-month old child	(0.187)	(0.267)	(0.252)	(0.369)
Larger housework share	-0.079	-0.370	0.282	-0.650*
5-year old child	(0.174)	(0.252)	(0.236)	(0.346)
Larger housework share	-0.391**	-0.358	-0.381	0.038
15-year old child	(0.184)	(0.259)	(0.262)	(0.368)
Age	-0.044***	-0.043***	-0.040**	-0.003
	(0.011)	(0.015)	(0.017)	(0.023)
Age square	0.000***	0.000***	0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	5.875***	5.933***	5.733***	0.205
	(0.302)	(0.427)	(0.433)	(0.608)
Number of observations	4,547	2,476	2,071	4,547
Number of individuals	1,541	841	700	1,541
Wald test chi squared	425.594	312.337	158.106	479.932
p-value	0.000	0.000	0.000	0.000
Rho	0.473	0.434	0.518	0.472

Notes: Asterisks denote significance levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(2) Model with equal coefficients across gender and using the pooled sample of women and men,

(2) Separate model for women, (3) separate model for men,

(4) Model with different coefficients across gender and using the pooled sample of women and men.

The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(31)$ for Columns 1-3 and $\chi^2(63)$ for Column 4.

Rho is the fraction of the variance of the unobserved component explained by the random effect.

TABLE A2 ORDERED PROBIT MODEL FOR SATISFACTION WITH RANDOM INDIVIDUAL EFFECTS AND STANDARD ERRORS

Variables	(1) All	(2) Women	(3) Men	
Lower housework share	-0.925*** (0.158)	-0.973*** (0.216)	-0.898*** (0.234)	-0.169 (0.317)
Large housework share	-1.270*** (0.157)	-1.355*** (0.218)	-1.217*** (0.229)	-0.258 (0.314)
Both partners work part-time (Both PT)	-0.149 (0.124)	-0.178 (0.167)	-0.0798 (0.186)	-0.106 (0.25)
Respondent work full-time, partner part-time (FT-PT)	-0.621*** (0.12)	-0.843*** (0.159)	-0.348 (0.184)	-0.554* (0.242)
Respondent work part-time, partner full-time (PT-FT)	-0.266* (0.123)	-0.308 (0.168)	-0.183 (0.183)	-0.135 (0.248)
Respondent earns half (Earns half)	-0.0351 (0.105)	-0.0027 (0.141)	-0.0566 (0.158)	0.041 (0.211)
Respondent earns twice as much (Earns twice)	-0.136 (0.106)	-0.0895 (0.14)	-0.151 (0.162)	0.0535 (0.213)
Pay for housework	-0.00456 (0.0856)	-0.0395 (0.115)	-0.00475 (0.13)	-0.0467 (0.173)
6-month old child	-0.0239 (0.12)	0.102 (0.16)	-0.233 (0.182)	0.325 (0.241)
5-year old child	0.0127 (0.118)	0.247 (0.16)	-0.269 (0.177)	0.517* (0.238)
15-year old child	0.276* (0.126)	0.314 (0.169)	0.217 (0.191)	0.117 (0.254)
Lower housework share Both PT	0.262 (0.147)	0.3 (0.198)	0.19 (0.222)	0.131 (0.297)
Lower housework share FT-PT	1.021*** (0.146)	1.210*** (0.194)	0.808*** (0.225)	0.52 (0.294)
Lower housework share PT-FT	0.456** (0.148)	0.383 (0.201)	0.507* (0.219)	-0.0845 (0.296)
Lower housework share Earns half	0.0185 (0.127)	-0.0427 (0.17)	0.0694 (0.191)	-0.11 (0.255)
Lower housework share Earns twice	0.367** (0.128)	0.366* (0.171)	0.313 (0.195)	0.0697 (0.259)
Lower housework share Pay for housework	0.0677 (0.104)	0.0837 (0.139)	0.0893 (0.157)	0.00446 (0.209)
Lower housework share 6-month old child	0.0111 (0.145)	-0.0151 (0.194)	0.12 (0.219)	-0.104 (0.291)
Lower housework share 5-year old child	-0.0817 (0.143)	-0.279 (0.193)	0.184 (0.215)	-0.468 (0.289)
Lower housework share 15-year old child	-0.478** (0.148)	-0.461* (0.198)	-0.470* (0.223)	-0.0211 (0.297)
Larger housework share Both PT	0.0936 (0.147)	0.0397 (0.199)	0.158 (0.221)	-0.112 (0.297)
Larger housework share FT-PT	0.457** (0.142)	0.638*** (0.19)	0.246 (0.218)	0.42 (0.288)
Larger housework share PT-FT	0.860*** (0.148)	1.013*** (0.203)	0.658** (0.219)	0.406 (0.298)
Larger housework share Earns half	0.135 (0.125)	0.0544 (0.167)	0.243 (0.189)	-0.166 (0.252)
Larger housework share Earns twice	0.0944 (0.131)	0.0127 (0.175)	0.166 (0.199)	-0.153 (0.264)
Larger housework share	0.268**	0.363**	0.181	0.216

Paid housework	(0.102)	(0.137)	(0.153)	(0.205)
Larger housework share	-0.00169	-0.116	0.229	-0.338
6-month old child	(0.147)	(0.198)	(0.222)	(0.297)
Larger housework share	-0.0612	-0.31	0.263	-0.565*
5-year old child	(0.14)	(0.191)	(0.21)	(0.283)
Larger housework share	-0.346*	-0.382	-0.275	-0.125
15-year old child	(0.153)	(0.205)	(0.231)	(0.308)
Age	-0.0371***	-0.0348**	-0.0359*	-0.00244
	(0.0099)	(0.0127)	(0.0159)	(0.0199)
	0.000415**			
Age square	*	0.000360**	0.000429**	-2.8E-05
	(0.0001)	(0.00013)	(0.00016)	(0.00021)
Observations	4,547	2,476	2,071	4,547
Number of individuals	1,541	841	700	1,541
Wald test Chi2	1006,78	440.52	570.49	1000,92
Wald p-value	0.000	0.000	0.000	0.000

Notes: Asterisks denote significance levels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(1) Model with equal coefficients across gender and using the pooled sample of women and men,

(2) Separate model for women, (3) separate model for men,

(4) Model with different coefficients across gender and using the pooled sample of women and men.

The Wald test relates to the joint significance of all coefficients in the column. It is distributed as $\chi^2(31)$ for Columns 1-3 and $\chi^2(63)$ for Column 4.

Rho is the fraction of the variance of the unobserved component explained by the random effect.