# Accepted Manuscript

Housework share between partners: Experimental evidence on gender-specific preferences

Katrin Auspurg, Maria Iacovou, Cheti Nicoletti

PII: S0049-089X(16)30222-8

DOI: 10.1016/j.ssresearch.2017.01.003

Reference: YSSRE 2034

To appear in: *Social Science Research*

Received Date: 21 April 2016

Revised Date: 2 January 2017

Accepted Date: 12 January 2017

Please cite this article as: Auspurg, K., Iacovou, M., Nicoletti, C., Housework share between partners: Experimental evidence on gender-specific preferences, *Social Science Research* (2017), doi: 10.1016/j.ssresearch.2017.01.003.

# HOUSEWORK SHARE BETWEEN PARTNERS:

# EXPERIMENTAL EVIDENCE ON GENDER-SPECIFIC PREFERENCES

KATRIN AUSPURG

*(Department of Sociology, LMU Munich)*[1]

MARIA IACOVOU

*(Department of Sociology, University of Cambridge)*[2]

CHETI NICOLETTI

*(DERS, University of York and ISER, University of Essex)*[3]

[1] (corresponding author) Konradstr. 6, DE-80801 Munich, Germany;
e-mail: Katrin.Auspurg@lmu.de
[2] Free School Lane, Cambridge CB2 3RQ, United Kingdom; e-mail: mi305@cam.ac.uk
[3] Heslington, York YO10 5DD, United Kingdom; e-mail: cheti.nicoletti@york.ac.uk

## Abstract

This paper uses a novel vignette-based experimental design to investigate the reasons underlying the gendered division of housework. We are particularly interested in the role of gender-specific preferences: are there differences in the utility that men and women derive from housework, and might these be responsible for the fact that women continue to do more housework than men? It is difficult to address these questions with conventional survey data, because of inherent problems with endogeneity and ex-post rationalization; our experimental design circumvents these problems. We find remarkably little evidence of any systematic gender differences in preferences, and a general inclination towards an equal distribution of housework; this suggests that the reasons for the gendered division of housework do not derive from gender differences in preferences, and must lie elsewhere.

**Keywords:** gender; housework; division of labor; factorial survey experiment; preference formation

# 1. INTRODUCTION

This paper reports the results of a novel experiment which investigates whether men and women have different preferences over the allocation of housework within couples, and asks whether such differences may provide an explanation for the fact that women do more housework than men.

Gender disparities in the allocation of housework have attracted interest 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; Thompson 1991; and many others). This issue is becoming ever more relevant in the contemporary Western context: women are now educated at least as well as men, the gender gap in labor market participation and earnings continues to narrow, but women continue to do much more housework than their male partners (Brines 1993, 1994; Álvarez and Miles 2003; Bitman et al. 2003; Washbrook 2007; Kan et al. 2011).

A range of theories have been advanced to explain what Hochschild (1989) refers to as the "stalled revolution." These theories will be discussed in Section 2; the debate essentially boils down to whether women do more housework because their capabilities, characteristics or standards 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 men's and women's preferences over housework differ systematically, with women liking housework more (or disliking it less) than men.

It is this hypothesis of systematically different gender preferences which this paper sets out to explore. There currently exists little empirical evidence on the gendered nature of preferences over the allocation of time, largely because of the difficulty of using survey data to obtain meaningful estimates of the relevant preferences. Although several household surveys carry questions on the allocation of paid work and housework between partners, and on individuals' satisfaction with these arrangements, the fact that these data relate only to people's *actual* arrangements (rather than what people would experience under *alternative* arrangements) leads to three main problems.

First, some distributions of housework and paid work are rarely observed in surveys (for example, surveys typically contain very few households where the woman does more paid work, earns more, and does less housework, than her male partner). 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 would be, given a range of possible alternatives. Pedulla and Thébaud (2015) show that preferences are sensitive to the institutional context, with people's preferences in the presence of institutional constraints such as a gender pay gap differing systematically from their "true" preferences in the absence of such constraints.

Third, people's hours of domestic and market work, as well as related factors such as their wages, are largely determined by their own characteristics and those of their partners – and these may be the same characteristics which drive their preferences over housework arrangements. Empirical analyses are thus subject to problems of endogeneity, meaning 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.[1]

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 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). Experimental studies on gender identity or gender-specific preferences include the laboratory experiment of Cadsby et al. (2013) investigating the effect of gender identity on attitudes to risk and competition; the experiment of Görges (2015) testing gender specific patterns in couples' work specialization decisions; the factorial survey experiment adopted by Abraham et al. (2010) testing the effect of gender role attitudes on migration decisions within dual-earner partners; and the survey-experiments run by Pedulla and Thébaud (2015) examining the extent to which institutional

[1] Hwang et al. (1998) demonstrate this point using simulation techniques, showing that estimates based on non-experimental data may represent individual's preferences very poorly; bias stemming from unobserved heterogeneity may even give rise to estimates with the wrong sign.

constraints, such as workplace policies, influence young, unmarried men's and women's preferences for their future work-family arrangements.

However, to best of our knowledge, no experimental research has yet been conducted to assess gender differences in preferences over housework arrangements. In our experiment people are invited to imagine themselves and their partners in different hypothetical domestic scenarios ("vignettes"), and to tell us how satisfied they would be with each set of arrangements. These hypothetical scenarios are generated using a multi-factorial experimental survey design (Auspurg and Hinz 2015)[2] which, as well as varying the distribution of housework between scenarios, also varies 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; and whether the household employs paid help (i.e. whether there is some market substitution of domestic work). Factorial survey methods are a proven method in the study of intragroup differences in judgement rules, such as cross-gender differences (Jasso 1994) and have been shown to be a useful tool for researching how individuals' preferences would change over a range of scenarios (e.g., Shlay 2010). In our design 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 preferences over work arrangements differ systematically between men and women, free from the problems of endogeneity and post-hoc rationalization. A finding of systematic differences in preferences would lead us to conclude that gender identity, i.e. the internalization of social gender norms, is a factor contributing to the unequal distribution of housework; conversely, a finding of few or no differences would lead us to conclude that the gendered division of housework arises not because of gendered 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 themselves, but which are enforced by some other means.

[2] Factorial survey experiments have been widely used by sociologists to study beliefs, attitudes and hypothetical decisions (see for a review: Wallander 2009). 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; Amaya-Amaya et al. 2008; Sándor and Franses 2009).

## 2. THEORETICAL BACKGROUND

The first formal model of choices relating to the allocation of housework and paid work between partners was proposed by Becker (1965), who hypothesized that asymmetric time allocations are due to 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 theoretical framework was originally proposed in terms of women having lower levels of human capital than men; however, it does not rely on this: gender asymmetries may arise from any 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. Additionally, specialization may also arise as the result of women being relatively more efficient than men in childcare and in other domestic activities: in this situation, women's comparative advantage in home-based production would mean that they end up doing more childcare and housework, even in a situation where both sexes were equally productive in the labor market.

Later social exchange and bargaining theories drop the very simplistic assumption of a joint utility function that underpins Becker's New Home economics, and hypothesize that individuals bargain over the allocation of housework, and follow both individualistic (maximization of *own* earning and bargaining power) and common preferences (maximization of joint household income). These models (for example, Ott 1992) predict that in a context where men have higher levels of human capital than women, men will do relatively more paid work (because of both their higher bargaining and earning power), while women do a larger share of the housework. However, they predict a partial rather than a full specialization, with both partners retaining some labor market power to preserve their economic independence and bargaining power within the partnership.

In theory this specialization is gender-neutral: if women and men had identical levels of human capital, we should observe both sexes doing similar shares of market work and housework, while if a man were comparatively more productive than his female partner in the domestic sphere, we would expect to observe him doing a larger share of the housework than

his partner.[3] However, several scholars have noted that women often do a larger share of housework even when (or in particular when) their market work share is as large as, or even larger than, their partner's (see, e.g. Akerlof and Kranton 2000; Schneider 2011; Greenstein 2000). This empirical evidence runs counter to the theoretical suggestion that work arrangements between partners are gender-neutral.

For this reason, sociologists have been critical of economic theories that assume gender neutrality, and have sought alternative explanations for the gender asymmetries observed in society. Hakim (2000) is the leading proponent of gender-specific preferences. In her "preference theory" she 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. However, Hakim primarily provides a description of different types of individuals, without explaining how and why the different preferences arise. Explanations proposed by other authors include women's lack of power within the family and in society at large (Lennon and Rosenfield 1994; Baxter and Western 1998); the moderation of gender values by social structures (welfare policies) that hinder couples in realizing egalitarian values, particularly once children are born (Bühlmann et al. 2009; Pedulla and Thébaud 2015) and social gender norms (Brines 1994; Baxter and Western 1998; Bianchi et al. 2000).

One promising explanation for gender-specific specialization is the existence of internalized gender norms. The internalization of gender norms occurs when people conform to the behavior and role prescribed by these norms in order to affirm their gender self-image (gender identity); behaviors and gender roles that deviate from those prescribed by social norms may cause anxiety and uneasiness and a loss of gender identity (West and Zimmerman 1987).[4,5] The economic model of Akerlof and Kranton (2000) integrates both the economic

[3] A gender-neutral specialization is also suggested by more recent economic papers (see Becker 1973; Gronau 1973 and 1977; Donni and Chiappori 2011).

[4] An alternative possibility to internalized norms would be external norms that are enforced by negative social sanctions in case of norm deviations or positive rewards in case of norm-compliance (Axelrod 1984; Ott 1992).

perspective on utility maximization, and the sociological perspective on internalized gender norms: individuals choose their working arrangements to maximize their utility, and utility depends on the consumption of goods and services made possible by these working arrangements, and also on the degree to which people's working arrangements reinforce their gender identity. Individuals whose work arrangements deviate from social gender norms will incur a penalty in utility compared to individuals whose choices affirm their gender self-image by conforming to customary gender roles.

A third strand of theory on domestic work arrangements, originating with Thompson (1991), applies the theory of distributive justice to the issue of housework, hypothesizing that satisfaction with domestic arrangements is related to individuals' perceptions of the fairness or justice of those arrangements (see Dette-Hagenmeyer and Reichle, 2016 for an overview[6]).

The literature on distributive justice defines several distinct principles of fairness (Deutsch, 1975), of which two are of particular relevance in the allocation of housework.[7] The principle of *equity* is fulfilled when both partner's rewards (for example leisure, or time not spent on housework) are distributed proportionally to individuals' inputs to the relationship. This leads to similar assumptions to those derived from exchange and bargaining theories: a family member is assumed to be entitled to less housework if he or she works longer hours or earns a larger salary (Thompson 1991; Gager 1998, 2008). The principle of *equality*, by contrast, defines as fair a situation in which rewards and duties are split equally between the partners, regardless of partners' inputs in the form of time spent at paid work or earnings; Jasso (1983) proposes that in small groups, and particularly in intimate relationships such as marital dyads, it is the principle of equality with regard to marital happiness which predominates; the empirical work of Gager (2008) also suggests that the principle of equality may predominate.

Without further development, the principles of both equity and equality would suggest gender-neutral allocations of housework; however, the empirical evidence suggests that both

---

However, in modern societies meaningful sanctions for people who deviate from a customary gender division of labor seem to be unlikely, and the same is true for formal rewards for those who conform.

[5] Carlson and Lynch (2013), using longitudinal data, suggest that the relationship might be the other way round: gender ideologies might change in order to avoid distress stemming from discrepancies between behavior and attitudes.

[6] "Interestingly, it is not so much the division of household labor itself, but rather the perception of it as just that exerts the main influence on well-being" (Dette-Hagenmeyer and Reichle 2016: 339).

[7] A third principle, that of need, is proposed along with equality and equity, but has more relevance to relations between adult and child family members, than to relations between adults.

women and men tend to perceive as "fair" a situation in which women do a disproportionate amount of domestic labour, and which is actually neither equal nor equitable (Major 1993; Lennon and Rosenfield 1994; Gager 1998, 2008). Drawing on the work of Major (1987), Thompson (1991) identifies three factors which may contribute to a sense of fairness with housework arrangements.

*Justification* relates to perceptions of the appropriateness of the procedures which led to the current allocation of work: for example, women may justify arrangements which may otherwise appear inequitable on the basis that their husbands' jobs are more important, or their husbands are not very good at housework, or are not able to multitask (Gager 1998).

*Comparison referent* relates to the benchmarks against which people compare their own arrangements. A woman doing the majority of housework in her home may feel her situation is unfair if she compares her contribution to that of her husband, but be more favourably disposed towards the situation if she compares her husband's contribution to the contributions of other husbands. Empirically, this hypothesis has received qualified support (Hawkins et al. 1995; Gager and Hohmann-Marriott 2006); however, the question remains as to why people should make between- rather than within-gender comparisons. One explanation is that they do this to reduce cognitive dissonance: individuals who find themselves with allocations they cannot easily escape might simply re-interpret their situation by switching the point of reference (Lennon and Rosenfield 1994; Gager 1998, 2008).[8] Alternatively, the gender roles, norms, identities and preferences that were discussed above, are likely to be key influences on people's choice of reference point.

*Outcome values* relate to the values people place on the outcomes of arrangements. Individuals must want an output in order to feel inequitably treated if they do not get it (Major 1987); women might consider it "fair" to do most of the housework, simply because they like doing family work more than men, because they appreciate outcomes such as family harmony more, or because they have higher standards for the housework than their male partners. Again, rationalizations for these gendered preferences typically rely on the theories of gender role socialization and gender identities discussed above.

[8] If a women compares her situation with that of a female friend, "she may experience herself as fortunate, as 'one of the lucky ones'" (Hochschild 1989; Dette-Hagenmeyer Reichle 2016: 337).

In this section, we have discussed three strands of theory, emanating from different disciplines in the social sciences, but all directed to the same question: why do women do so much more housework than men, even when they put in similar hours of paid work than men, and have similar earnings? Each of these theoretical approaches suggests different mechanisms which might be responsible for the gendered nature of housework; experimental evidence on whether preferences differ between men and women would make a contribution to all three.

Formally, we hypothesize that differences between the gender identities of men and women manifest themselves in differences in preferences over the allocation of domestic work, with gender norms driving women to like housework more (or dislike it less) than men. In the experiment which forms the empirical basis for our research, this hypothesis predicts that women would react more favourably than men to scenarios in which they do the majority of the housework.

## 3. STATE OF RESEARCH

As mentioned earlier, we are not aware of any existing research which directly examines the relationship between housework sharing within a partnership and the utility (e.g. measured by satisfaction or preferences) 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.

Kalmijn and Monden (2011) find a lower incidence of depressive symptoms among both husbands and wives in couples where housework is shared more equally. Lennon and Rosenfield (1994) find that depression is more common among women who feel that their housework arrangements are unfair; Harryson et al. (2012) find 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.

Several studies have sought to assess whether the principles of distributive justice explain perceptions of fairness with housework. Gager (1998, 2008), when interviewing 25 dual earner couples with children in the US, found evidence that most spouses support the principles of both equality and equity, with earnings and the time spent on paid labor being particularly relevant considerations of equity. Nevertheless, many couples ended up with women performing more housework but not reporting their situation as being unfair, leading the author to conclude that a range of justifications are used to mitigate the conflict with individuals' true sense of fairness. Gager and Hohmann-Marriott (2006) analyzed quantitative data from a national survey on housework in the US, concluding that the best predictor of individuals' evaluations of fairness was a combination of gender role norms and marital equity in the form of time inputs in paid and unpaid labor. Hawkins et al. (1995) find the distributive justice framework to be effective at explaining the perceived fairness of housework arrangements, and suggest that joint decision-making by couples, and familial appreciation of domestic work, are also important determinants of perceptions.

A sizeable literature examines 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. Similarly Brines (1993, 1994) and Greenstein (2000) found non-linear relationships between women's

share of household income and amount of housework: women earning less but also women earning more than their partner did more than half of the housework. These results have been interpreted as evidence for the gender-display thesis: large amounts of housework might allow female "breadwinners" to compensate for the deviance from their gender role. Gupta (2007) suggests that this pattern might not be due to women's share of earnings, but by their absolute levels of earnings (which are correlated with their share of earnings); the higher women's absolute earnings, the more likely they are to engage paid help, leading to a decrease in their hours of housework. This implies a gender-specific process of buying out of housework for which the underlying mechanisms sill have to be explored (and see Schneider, 2011 for contrasting evidence supporting the gender-display thesis, even when partners' absolute earnings are controlled). 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 analyze 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.

In summary, the evidence suggests that the distribution of housework in couples is in many cases neither equal nor equitable, with women doing disproportionate shares of housework even when their hours of paid work and wages are taken into account. It suggests that both men and women are motivated by considerations of fairness, but that these considerations may not be reflected in actual allocations of work. This suggest a degree of ex-post rationalization (which, as we discussed in the Introduction, is one of the issues which means it is difficult to assess preferences with survey-based research, and one of the reasons we use an experimental approach). The only other experimental work of which we are aware, which deals with household allocation, is that of Pedulla and Thébaud (2015), who conducted an experiment using a probability-based online panel of young, unmarried, childless people in the US, asking respondents how they would like to balance work and family responsibilities

in their future relationships, under a range of hypothetical institutional constraints (in particular, whether workplaces provided paid family leave, subsidized childcare, and flexible work options). The main finding was a clear preference for equal sharing of paid and unpaid work: when respondents could opt for this type of relationship, a clear majority of both men and women chose this option; and in particular women's preferences were responsive to supportive work-family policy interventions. However, the authors only tested a joint bundle of arrangements (not varying the amount of paid and unpaid work or supportive policies separately), and the between-respondent analytical design in this study means that it cannot examine how given individuals might shift their preferences under different sets of constraints. The current research builds on this work, using a within-respondent variation and more fine-grained scenarios describing possible work-family arrangements.

## 4. 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 (IP5), 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 seven-point scale ranging from *completely dissatisfied* to *completely satisfied* (see Figure 1).

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 | - | - | - |

*Note*: Resp. stands for respondent.

The full set of possible scenarios spans all 480 possible combinations[9] 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 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 |
|---|---|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

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).[10]

[9] The number of possible combinations is the product of the number of categories:
$n = 4x3x4x5x2 = 480$.

[10] The experimental factors were built up in order to maximize the variance of the different levels; this provides the maximum statistical power for estimating the relationship between the vignette dimensions and the dependent variable (in our case, satisfaction with housework distributions). This means that the distributions of the factors in the vignette scenarios are not the same as the distributions that occur in reality. For instance, in our "vignette world", about 50 percent of vignettes specify that the couple employs a cleaner, while in real life, the proportion employing a cleaner would be much smaller. Likewise, very few women earn double what their partners earn, but this is the scenario given to one third of female respondents (see also the descriptive statistics in Table 2). The ability to include such scenarios in numbers which do not correspond to reality is an important strength of the experimental design, since it provides weight in parts of the joint distribution which in real life are sparsely populated.

These 160 questionnaire versions were randomly allocated to households participating in the experiment,[11] 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.

We also included the following preamble before the three vignette questions: "We are interested in the way people feel about how couples share housework. You will be asked about three different scenarios, each describing an imaginary situation relating to different work arrangements between couples. In some of the situations the couple may have children. In these cases you should assume that both partners are involved in their day-to-day care, and that both are happy with the amount of time they spend with their children. We are really only interested in knowing how you feel about the way housework is shared. There are no right or wrong answers!"

The experiment was administered in self-completion mode via computer-assisted self-interview (CASI).[12] 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

[11] 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).

[12] 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.

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; Sonnenberg et al. 2012). 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 descriptions of scenarios. Thus, all gender differences in earning power and the shares of housework and market work have been levelled 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.

It is worth noting that although measures of self-reported satisfaction are now widely used in social science research, and are widely considered to be valid indicators of wellbeing (Diener et al. 1999; Kahneman and Krueger 2006; Clark et al. 2008), they have not been accepted with universal enthusiasm, and some social scientists remain skeptical about their reliability (see Bertrand and Mullainathan 2001, and Ravallion and Lokshin 2001 for a review). One concern is that individuals might differ in their use of the scale provided to rank their situation (Kristensen and Johansson 2008). This problem is addressed in the current study by the fact that we collect multiple observations on participants and we control for individual effects, showing that unobserved individual-specific characteristics (such as personality traits and response styles) are uncorrelated with the experimental factors explaining levels of reported satisfaction.

### *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

experiments.[13] Some of these evaluated only one or two of the three vignettes; thus, a total of 4,547 valid evaluations were generated. This means that 66% of adults in participating households provided responses to at least one vignette, and 62% of the vignettes administered to adults in participating households yielded responses. 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. The first figures relate to the dependent variable (satisfaction with housework arrangements). We calculate a mean value for this variable by assigning the numeric values 1-7 to the response categories (so *completely dissatisfied* is associated with the value 1, while *completely satisfied* is associated with the value 7). The mean of the variable is just over 4, the midpoint of the range; the average satisfaction rating is slightly higher for men than for women.

[13] 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

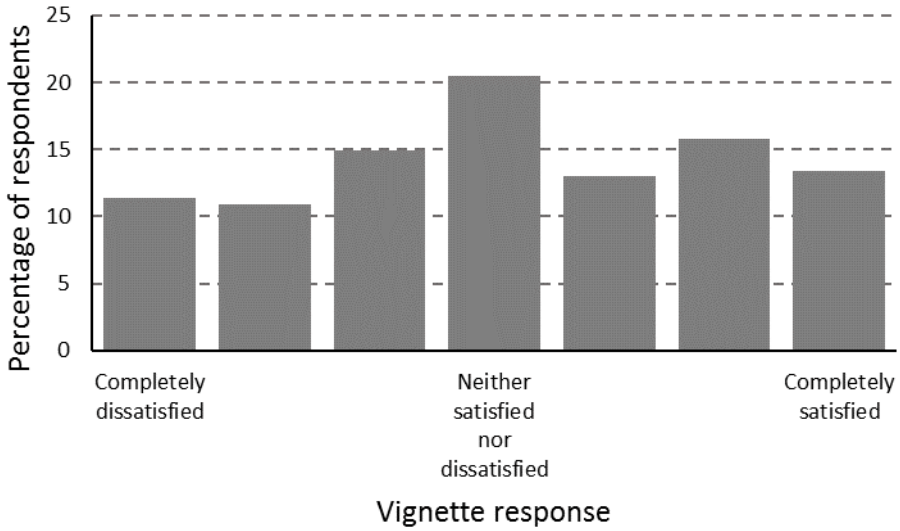
| | Mean | Standard deviation |
|---|---|---|
| Dependent variable: self-reported satisfaction | | |
| Women | 4.042 | 1.919 |
| Men | 4.261 | 1.860 |
| All | 4,142 | 1.895 |
| Descriptive statistics relating to vignette dimensions and categories | | |
| Share of housework (ref: respondent does about 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 pay (ref: resp. and partner's pay about equal) | | |
| Respondent's pay half that of partner | 0.343 | |
| Respondent's pay double that of partner | 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 of respondents | | |
| Male | 0.455 | |
| Age | 47.711 | 16.845 |
| Married | 0.576 | |
| 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 | |
| Vocational/professional and other qualifications | 0.189 | |
| University degree | 0.280 | |
| Number of individuals | 1,609 | |
| Number of observations | 4,547 | |

*Notes*: The statistics in the second panel (statistics relating to vignette dimensions and categories) relate to the characteristics of the scenarios administered to respondents, and *not* to the characteristics of respondents themselves.
Educational categories relate to the UK qualification system. GCSEs/O-levels are school-based qualifications obtained via examination at age 16; A-levels are academic qualifications obtained via examination at age 18. Individuals with both a university degree and vocational/professional qualifications appear in the "university degree" category.

We then present 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.

The real-life characteristics of participants are shown in the lower part of Table 2. Around half the sample is male, the average age of respondents is around 48 years, and around 70% are actually married or living with a partner. About a quarter of our respondents has a university degree, which matches the actual share in the UK population fairly well.

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, 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 |r| = .04, demonstrating that condition (a) is satisfied. All cross-correlations between the factors varying between questions were also well below .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:

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

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; $\mu_i$ 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 $\varepsilon_{is}$ 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 clustered robust standard errors to take account of the possible correlation in the error term within households. 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[14]; we also control for individual age and age squared.[15]

*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. 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 gender.[16] 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. 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

[14] This yields a model with 18 interaction terms. Models with large numbers of interaction terms may suffer from problems of low power, driven by multicollinearity between regressors (Friedrich 1982). In our analysis this is unlikely to be a substantial problem because the experimental design minimizes correlation between experimental factors. However, we report results from alternative specifications with smaller numbers of interaction terms in Section 4.3.

[15] Additional controls for the respondents' earnings level, actual family and employment status, our housework share did not change the substantive results. Given the experimental design, we used parsimonious models restricted mostly to experimental stimuli.

[16] 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.

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 that as well as demonstrating an overall preference for equality, both men and women are also motivated by considerations of equity in the total allocation of work.

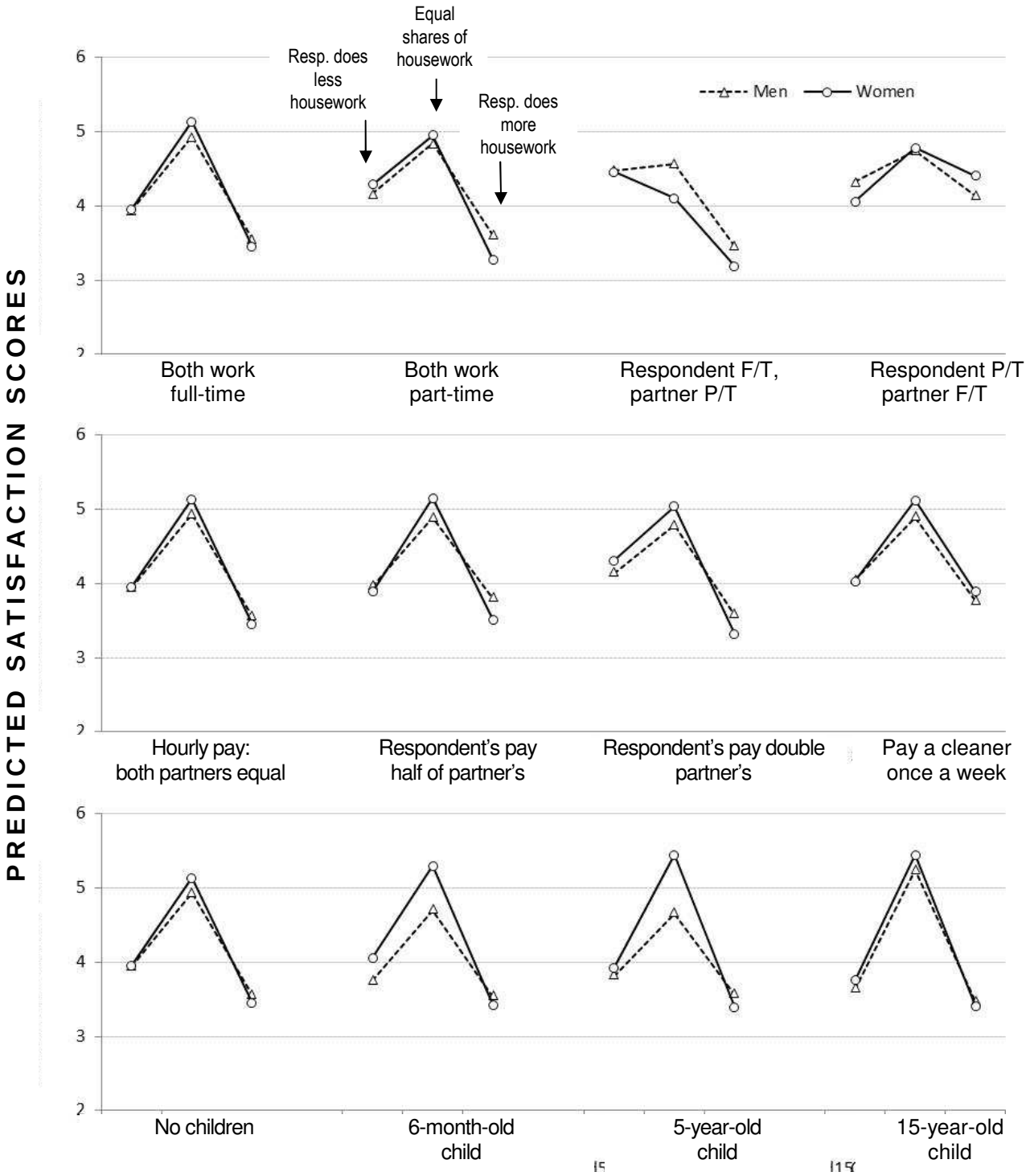
The coefficients in Table 3 are most easily interpreted graphically. Figure 3 shows predicted satisfaction levels for men and women, varying the vignette factors.

TABLE 3: LINEAR RANDOM EFFECTS MODEL OF SATISFACTION RATINGS

| | (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 | -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 pay (ref: about equal) | | | | |
| Resp. hourly pay half that of partner | -0.033 | 0.015 | -0.060 | 0.063 |
| Resp. hourly pay double that of partner | -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 pay half that of partner | 0.009 | -0.082 | 0.077 | -0.154 |
| Resp. hourly pay double that of partner | 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 pay half that of partner | 0.167 | 0.049 | 0.299 | -0.244 |
| Resp. hourly pay double that of partner | 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 | 0.0005** | 0.0005** | 0.0005** | -0.0003 |
| 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 $\chi^2$ | 425.59 | 312.34 | 158.11 | 479.93 |
| *p*-value | .000 | .000 | .000 | .000 |
| Rho | 0.473 | 0.434 | 0.518 | 0.472 |

*Notes*: Asterisks denote significance levels. ** $p < .01$, * $p < .05$. Standard errors are robust to correlation between individuals belonging to the same household.
(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 faction of the variance of the unobserved component explained by the random effect.

FIGURE 3: PREDICTED MEN AND WOMEN 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; $n$ = 4,547 vignette judgements and $n$ = 1,541 respondents.

Predicted values are calculated at the midpoint of the age variable (about 48 years) and, unless stated, for the reference groups for all the vignette variables (which are the labor market categories “both full-time;” and “hourly pay about equal;” and for the further categories: “no cleaner;” “no children”). Predictions are estimated using the pooled sample of men and women, imposing a zero random effect and a zero idiosyncratic error $\varepsilon_{is}$.

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. Confidence intervals have not been shown on the graph (they are available on request), but the gender differences fall well short of statistical significance, even at the 10% level.

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. We find no gender difference in the level of satisfaction when respondents consider a scenario where they work full-time, their partner works part-time, and the respondent does less housework than their partner. However, women are less happy than men in situations where they work full-time, their partner works part-time, and they do the same amount of housework, or more, than their partner (top panel of the Figure 3). In other words, women seem to have a stronger preference for adjusting their housework share downwards when they do the majority of paid work. This asymmetry is also evident in the scenarios where the respondent works part-time and the partner works full-time; in this scenario, women prefer doing more housework than their partner over doing less, while men seem indifferent between doing more or less housework. These results are congruent with the empirical evidence produced by Gough and

Killewald (2001), who find that people increase their share of housework on losing their job, but women increase twice as much their housework share than men.[17]

Note that in scenarios where the woman works full-time, while the man works part-time, women prefer a share of housework that is less than half, while men prefer an equal share. In other words, scenarios where a woman works full-time while the man works part-time may lead to some disagreements on how to allocate the share of housework between partners. If the allocation of housework is decided by maximizing the sum of the satisfaction of the two partners, then women who work full-time while their partner works part-time would end up doing as much housework as their partner. This obviously decreases their level of satisfaction and may cause conflict and disagreements within partners, which could be resolved by avoiding arrangements where women do more paid work than their partner.

In virtually all other scenarios, it is clear that the differences between predicted values for men and women are very small. There are two exceptions: when a 6-month-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.[18] It is interesting that, even in scenarios in which (small) children are present (i.e. those scenarios where in reality very traditional divisions of labor might be observed), respondents tended to opt for equal allocations where these were available. This is in line with the results of Bühlmann et al. (2009) and Pedulla and Thébaud (2015): in the absence of (institutional) constraints, individuals prefer gender neutral over gendered arrangements.

Testing formally for gender differences across the full model allowing all coefficients to differ between men and women, a Wald test (distributed as $\chi^2$ with 32 degrees of freedom) takes the value 46.77, with a *p*-value of .045. Thus, at the 1% level of significance, we do not reject the hypothesis that all regression coefficients between genders are equal, but at 5% level we do.

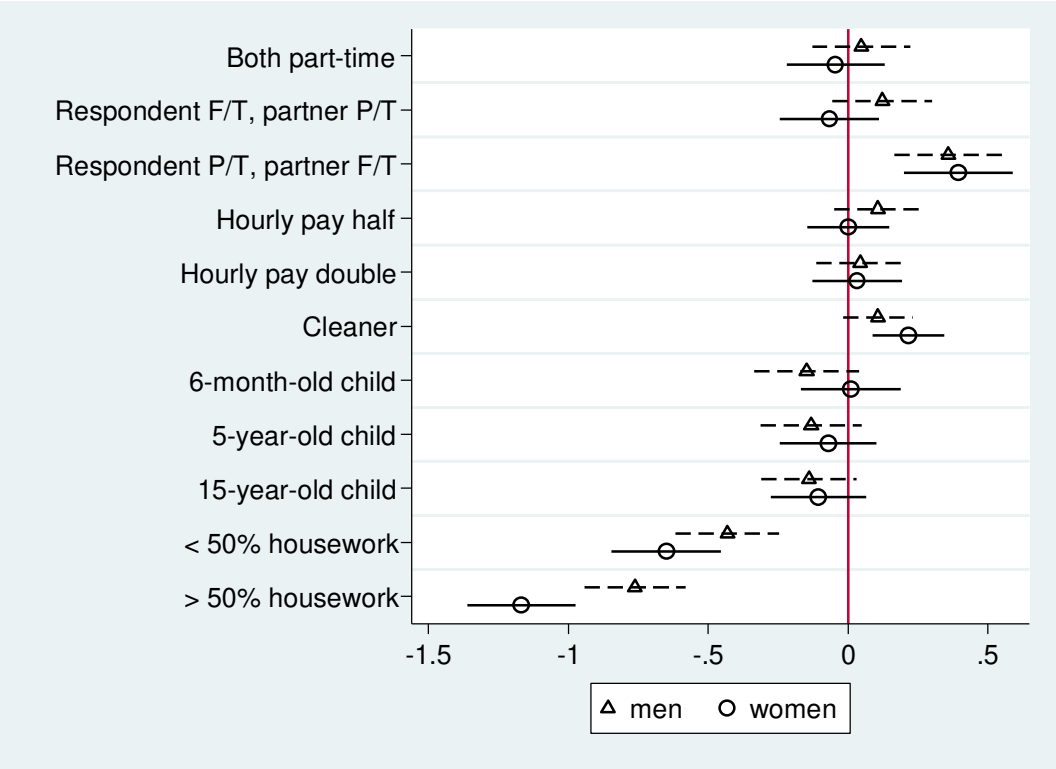
[17] At the same time this results runs counter the assumption often found in justice literature, that women have a stronger preference for equality (see, e.g., Major 1987, 1993). However, also other more recent work on earnings inequalities did not find support for such gender differences in justice principles (Auspurg et al., forthcoming).

[18] In addition one can see from Figure 3 that following respondents' evaluations the time budget is more relevant than their earning power in determining ideal shares of housework: both men and women prefer to exchange higher working hours against lower housework shares, while having a relatively higher hourly pay has *ceteris paribus* (given that both perform an equal amount of paid work—both work full-time) only a small exchange value in regard to an ideal share of housework.

Although the structures of men's and women's preferences appear to be similar, it may be that these similarities relate only to mean effects, and that they mask a degree of heterogeneity. In particular, preference theory (Hakim 1996, 2000) which suggests that women are much more heterogeneous than men in their preferences over the allocation of work in the household, might predict a higher degree of dispersion in our estimates for women. Figure 4 plots estimated coefficients from the regressions reported in Table 3, together with 95% confidence intervals. The variance in mean satisfaction ratings, measured by the standard deviation of the respondent-specific random intercept $\mu_i$ , is actually slightly larger for men than for women;[19] and there is no evidence that women's preferences over the different employment or housework options presented in the vignettes have a higher variance (larger confidence intervals) than those of men; women generally show a strong tendency to avoid very home-centered arrangements where they have to fulfill a larger share of housework than their partner.[20]

[19] Running separate random intercept regressions for male and female respondents, the estimated standard deviations of random intercepts are: sd($\mu_i$_men) = 1.13; sd($\mu_i$_women) = 0.99.
[20] Fagan (2001) and McRae (2003) also test (and reject) preference theory, with different methodologies.

FIGURE 4: EVALUATION OF VIGNETTE DIMENSIONS BY GENDER OF RESPONDENT




*Notes:* The figure shows the β-coefficients of the different vignette dimensions attached with 95% confidence intervals. The coefficients result of a random intercept regression with $n$ = 2,476 vignette judgments of $n$ = 841 female respondents and $n$ = 2,071 vignette judgments of $n$ = 700 male respondents. Reference categories: both full-time; earnings (hourly pay): both partners equal; no cleaner; no child; equal shares of housework (50%).

All in all 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 (which supports the findings of both Bühlmann et al. 2009 and Pedulla and Thébaud 2015) is of a preference for equality: 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, suggesting that couples invoke considerations of equity as well as equality. Both are gender-neutral principles, which 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 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 are used. In this section, we deal with a number of issues relating to the validity of our experiment and of the associated results.

*Non-linearities in the scale of reported satisfaction*

The linear model presented in Table 3 implicitly assumes that the 7-point scale of reported satisfaction is a direct measure of utility, and that each increment on the 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, 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 the Appendix).

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 pay (ref: about equal) | | | | |
| Resp. hourly pay half that of partner | -0.035 | -0.003 | -0.057 | 0.041 |
| Resp. hourly pay double that of partner | -0.136 | -0.089 | -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 pay half that of partner | 0.019 | -0.043 | 0.070 | -0.110 |
| Resp. hourly pay double that of partner | 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 pay half that of partner | 0.135 | 0.054 | 0.243 | -0.166 |
| Resp. hourly pay double that of partner | 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** | -0.036* | -0.002 |
| Age squared | 0.0004** | 0.0004** | 0.0004** | -0.00003 |
| Observations | 4,547 | 2,476 | 2,071 | 4,547 |
| Number of individuals | 1,541 | 841 | 700 | 1,541 |
| Wald test $\chi^2$ | 397.52 | 299.79 | 158.49 | 447.02 |
| Wald *p*-value | .000 | .000 | .000 | .000 |
| Rho | 0.523 | 0.498 | 0.553 | 0.524 |

*Notes*: Asterisks denote significance levels. ** $p < .01$, * $p < .05$. Standard errors are robust to correlation between individuals belonging to the same household.
(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 faction of the variance of the unobserved component explained by the random effect.

*The number of interaction terms*

The model presented in Table 3 includes 18 interaction terms. A known feature of models with large numbers of interaction terms is that they are susceptible to problems of low power, because of multicollinearity between regressors (Friedrich 1982). This is unlikely to be a significant problem in the context of the current analysis, since the experimental design minimizes correlation between the experimental factors, and therefore between the different interaction terms. Nevertheless, it is worth investigating whether the lack of significant differences between men's and women's preferences could arise from the inclusion of large numbers of regressors. We therefore re-estimate five variants of the model, including only interactions with (a) hours of work variables; (b) hourly pay variables; (c) the help with housework variable (cleaner); (d) the presence of children variables; and (e) no interaction terms at all. All of these models yield results which are qualitatively similar to those presented in Table 3; in particular, all share the finding of a significant preference among both sexes for an equal distribution of housework, and a general absence of differences between the coefficients for men and women. However, one interesting finding does emerge. Table 3 revealed that the general distaste for doing a larger share of the housework was slightly more pronounced among women than among men, but this was not statistically significant. In three of the reduced models (a, c and e) this difference becomes statistically significant at the 5% level. This runs counter to our hypothesis: if preferences were driven by internalized social norms, we would expect women to display less, not more, distaste for a situation in which they do the majority of the housework.

*Random distribution of vignettes*

Although vignettes (and factors within vignettes) were randomly assigned to sample members, it does not necessarily follow that vignettes are completely randomly distributed among *respondents*, since some people do not respond to one or more of the questions. We investigate this possibility 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 re-estimating the model 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 uncorrelated with individual-level unobservable characteristics. The computed *p*-value of this test is .375; 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 model, including as additional explanatory variables individuals' real-life levels 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.

*Comparability of responses between individuals*

A further potential 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 (for example, because responses are influenced by personal perceptions and emotional factors such as a respondent's state of mind at the time of interview). This potential problem is addressed by the same comparison of fixed and random effects coefficients that was used above. Fixed effects estimates "net out" any effects arising from between-individual differences, while random effects models do not; 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.[21]

*Variations between population subgroups*

In our experiment we assume that people abstract from their real-life situation when rating hypothetical scenarios. But this assumption does not automatically hold: factors such as the presence or absence of a partner or children, or real-life housework arrangements, may influence the relationship between gender and preferences. Thus, although we observe few

[21] Of course, in a survey where people are asked about their satisfaction with their real-life arrangements, factors describing partners' work arrangements, earning power and so on would *not* generally be uncorrelated with other individual characteristics, and this could lead to biased estimates.

differences between men's and women's preferences across the whole sample, it is possible that gender differences may exist 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 more 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.

TABLE 5: TESTING GENDER DIFFERENCES IN THE COEFFICIENTS OF THE LINEAR REGRESSION MODEL WITH INDIVIDUAL RANDOM EFFECTS

| Sample | Test statistics | *p*-value | No. obs. |
|---|---|---|---|
| Full sample | 46.77 | .045 | 4,547 |
| Married people | 39.11 | .181 | 2,608 |
| Single people | 40.86 | .136 | 1,939 |
| People with children | 54.78 | .007 | 1,230 |
| People without children | 32.74 | .430 | 3,317 |
| Housework share ≤ 50% | 33.50 | .395 | 1,606 |
| Housework share > 50% | 44.43 | .071 | 2,941 |
| Gender primed | 26.23 | .753 | 2,334 |
| Excluding "Satisficers" | 35.08 | .324 | 3,081 |

*Notes*: The variables used to define the different subsample are real-life characteristics. The test is supposed to be distributed as a $\chi^2$ with 32 degrees of freedom under the assumption of gender equality in the coefficients. Estimations are robust to correlation between individuals belonging to the same household.

We report the results of these tests in Table 5. 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 .181 for people who are married or living with a partner, and .136 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 .430 for people without children and .007 for people with children. This gender difference among people with children, which is significant at 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 .395), whereas among people who actually do more than half the housework, the gender differences are not significant at the 10% but not at the 5% level (*p*-value .071). In this case, the differences are driven mainly by differences in the effects of a five-year-old child (associated with 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 generally reject the hypothesis that men's and women's preferences for work arrangements are identical in these subsamples.

*Abstraction from own gender*

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 respond as generic individuals rather than as members of their own gender. We carry out a sensitivity analysis based on a subsample of people for whom gender identity had been "primed" in the survey 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 re-estimate the model with individual random effects for this subsample, and test between-gender differences, we do not reject the equality of the coefficients between genders (the test result is reported in the penultimate row of Table 5). 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 answering the vignette questions.

*"Satisficing" by survey respondents*

"Satisficing" (Krosnick et al. 1996; Oppenheimer et al. 2009) is a practice whereby, when responding to a survey, some respondents provide answers to questions, but shortcut their cognitive efforts, so 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, often 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). Results are presented in the final row of Table 5; 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, 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. The only exceptions are the following: (1) in situations where there is an unequal sharing of paid work between partners, women have a slightly stronger preference then men for adjusting their housework share in response to their paid work share; (2) in scenarios where the woman works full time and the man works part-time, men's and women's preferences diverge, with women preferring to do less housework than their partner and men preferring to do an equal amount of housework; (3) in the presence of a young child (6-month- or 5-year-old) women's preference for an equal share of housework intensifies more than that of men. However, in the small number of cases where we observe differences, they are small, and actually go against the gender norm.

In terms of the theoretical frameworks which we discussed earlier, our findings provide little evidence for the sort of gendered preferences proposed by Hakim (2000) or by Akerlof and Kranton (2000). Our finding of a marked preference, among both men and women, for equal and/or equitable arrangements, provides strong support for the applicability of distributive justice theory to this area (Major 1987; Thompson 1991); the reasons for

enduring inequalities in the allocation of housework, even among a sample of individuals who express a preference for equal arrangements, remains an open question.

Our results are subject to the caveat that they are derived from an experiment conducted in a single country (the United Kingdom), and cannot necessarily be generalised to other countries. The distribution of housework varies between countries, as do factors relating to norms, culture and financial incentives. In a country such as Germany, where the allocation of housework and paid work is far more gendered than in the UK, and where this allocation is underpinned by the tax and benefit system associated with a conservative welfare state, the findings of a similar study may be very different.

That said, our results are congruent with results from studies based outside the UK which, while they do not directly measure preferences, suggest an underlying preference for a more equal distribution of housework. Harryson et al. (2012), using a Swedish sample, suggest that levels of psychological distress are lower, among both men and women, in partnerships in which housework is shared equally; Lennon and Rosenfield (1995) use a US sample, finding that depressive symptoms are more common among women who perceive their housework arrangments as unfair; Kalmijn and Monden (2011), also using a US sample, find a lower incidence of depressive symptoms among both husbands and wives in couples where housework is shared more equally.

Given our empirical results, our main conclusion is that (in the UK at least) women's preferences are not aligned with gender norms, so that the reasons behind the gendered allocation of housework must lie elsewhere, perhaps in the different bargaining strategies employed by men and women.

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. Our empirical evidence provides a potential explanation for this type of behavior in the scenario where women have a full-time work while their partner works part-time. In this scenario men are happier with an equal share of housework, while women are happier with a lower share of housework. This situation can create some conflict between partners, which might be resolved by women doing actually half of the housework or by women specializing more in housework and avoiding situations where they do more paid work than their partner.

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.

***Acknowledgements:*** *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 XXX and XXX. We are grateful to seminar participants at the Universities of Cambridge, Canterbury and York, and to anonymous referees at this journal.*

**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, and C. Sauer (forthcoming). Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments. Accepted for publication in *American Sociological Review*.

Auspurg, K., T. Hinz, C. Sauer, and S. Liebig (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.

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(299): 493-517.

Becker, G.S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy* 81(4): 813-846.

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.

Browning, M., P.A. Chiappori, and V. Lechene (2010). Distributional Effects in Household Models: Separate Spheres and Income Pooling. *The Economic Journal* 120(545): 786-799.

Bryan, M.L. and A. Sevilla-Sanz (2011). Does Housework Lower Wages? Evidence for Britain. *Oxford Economic Papers* 63(1): 187-210.

Buck, N. and S. McFall (2012). Understanding Society: Design Overview. *Longitudinal and Life Course Studies* 3(1): 5-17.

Bühlmann, F., G. Elcheroth, G., and M. Tettamanti (2009). The Division of Labour Among European Couples: The Effects of Life Course and Welfare Policy on Value-Practice Configurations. *European Sociological Review* 46(1): 49-66.

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., M. Servatka, and F. Song (2013). How Competitive are Female Professionals? A Tale of Identity and Conflict. *Journal of Economic Behavior & Organization* 92(C): 284-303.

Carlson, D.L., and J.L. Lynch (2013). Housework: Cause and Consequence of Gender Ideology? *Social Science Research* 42(6): 1505-1518.

Clark, A.E., P. Frijters, and M.S. Shields (2008). Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature* 46(1): 95-144.

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.

Dette-Hagenmeyer, D., and B. Reichle (2016). Justice in the Couple and the Family, pp. 333-348 in C. Sabbagh, and M. Schmitt (eds.), *Handbook of Social Justice Theory and Research*. New York: Springer.

Deutsch, M. (1975). Equity, Equality and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice? *Journal of Social Issues* 31(3): 137-149.

Diener, E., E.M. Suh, R.E. Lucas, and H.L. Smith (1999). Subjective Well-Being: Three Decades of Progress. *Psychological Bulletin* 125(2): 276-302.

Donni, O., and P.A. Chiappori (2011). Non-Unitary Models of Household Behavior: A Survey of the Literature, pp. 1-40 in J. Alberto Molina (ed.) *Household Economic Behaviors*. New York: Springer.

Fagan, C. (2001). Time, Money and the Gender Order: Work Orientations and Working-Time Preferences in Britain. *Gender, Work and Organization* 8(3): 239-266.

Friedrich, Robert. 1982. In Defense of Multiplicative Terms in Multiple Regression Equations. *American Journal of Political Science* 26(4):797-833.

Gager, C.T. (1998). The Role of Valued Outcomes, Justifications, and Comparison Referents in Perceptions of Fairness Among Dual-Earner Couples. *Journal of Family Issues* 19(5): 622-648.

Gager, C.T. (2008). What's Fair is Fair? Role of Justice in Family Labor Allocation Decisions. *Marriage & Family Review* 44(4): 511-545.

Gager, C.T., and B. Hohmann-Mariott (2006). Distributive Justice in the Household. *Marriage & Family Review* 40(2-3): 5-42.

Greenstein, T.N. (2000). Economic Dependence, Gender, and the Division of Labor in the Home: A Replication and Extension. *Journal of Marriage and the Family* 62(2): 322-335.

Görges L. (2015). The Power of Love: A Subtle Driving Force for Unegalitarian Labor Division? *Review of the Economics of the Household* 13(1): 163-192.

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.

Gronau, R. (1973). The Intrafamily Allocation of Time: The Value of the Housewives' Time. *The American Economic Review* 63(4): 634-651.

Gronau, R. (1977). Leisure Time, Home Production, and Work. The Theory of Allocation Revisited. *Journal of Political Economy* 85(6): 1099-1123.

Gupta, S. (2007). Autonomy, Dependence, or Display? The Relationship between Married Women's Earnings and Housework. *Journal of Marriage and Family* 69(2): 399-417.

Hakim, C. (1996). The Sexual Division of Labor 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.

Hawkins, A.J., C.M. Marshall, and K.M. Meiners (1995). Exploring Wives' Sense of Fairness About Family Work. An Initial Test of the Distributive Justice Framework. *Journal of Family Issues 16*(6): 693-721.

Hochschild, A (1989). *The Second Shift: Working Parents and the Revolution at Home*. New York: Viking.

Hersch, J., and L.S. Stratton (1994). Housework, Wages, and the Division of Housework Time for Employed Spouses. *American Economic Review* 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.

Hwang, H., D.T. Mortensen, and R. Reed (1998). Hedonic Wages and Labor Market Search. *Journal of Labor Economics* 16(4): 815-847.

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.

Jasso, G. (1983). Social Consequences of the Sense of Distributive Justice: Small-group Applications, pp. 243-294 in D.M. Messick and K.S. Cook (eds.), *Theories of Equity: Psychological and Sociological Perspectives*. New York: Praeger.

Jasso, G. (1994). Assessing Individual and Group Differences in the Sense of Justice: Framework and Application to Gender Differences in the Justice of Earnings. *Social Science Research* 23(4): 368-406.

Kahneman, D., and Krueger, A.B. (2006). Developments in the Measurement of Subjective Well-Being. *Journal of Economic Perspectives* 20(1): 3-24.

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.

Kalmijn, M., and C.W. Monden (2011). The Division of Labor and Depressive Symptoms at the Couple Level: Effects of Equity or Specialization? *Journal of Social and Personal Relationships* 29(3): 358-374.

Kristensen, N., and E. Johansson (2008). New Evidence on Cross-Country Differences in Job Satisfaction using Anchoring Vignettes. *Labour Economics* 15(1): 96-117.

Krosnick, J.A., S. Narayan, and W.R. Smith (1996). Satisficing in Surveys: Initial Evidence. *New Directions for Evaluation* 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.

Major, B. (1987). Gender, Justice and the Psychology of Entitlement, pp. 124-148 in P. Shaver and C. Hendricks (eds.), *Sex and Gender. Review of Personality and Social Psychology* (Vol. 7). Newbury Park, CA: Sage.

Major, B. (1993). Gender, Entitlement, and the Distribution of Family Labor. *Journal of Social Issues* 49(3): 141-159.

McRae, S. (2003). Constraints and Choices in Mothers' Employment Careers: A Consideration of Hakim's Preference Theory. *British Journal of Sociology* 54(3): 317-388.

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.

Pedulla, D.S., and S. Thébaud (2015). Can We Finish the Revolution? Gender, Work-Family Ideals, and Institutional Constraint. *American Sociological Review* 80(1): 116-139.

Ravallion, M., and M. Lokshin (2001). Identifying Welfare Effects from Subjective Questions. *Economica* 68(271): 335–357.

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.

Schneider, D. (2011). Market Earnings and Household Work: New Tests of Gender Performance Theory. *Journal of Marriage and Family* 73(4): 845-860.

Shlay, A.B. (2010). African American, White and Hispanic Child Care Preferences. A Factorial Survey Analysis of Welfare. *Social Science Research* 39(1): 125-141.

Sigle-Rushton, W. (2010). Men's Unpaid Work and Divorce: Reassessing Specialization and Trade in British Families. *Feminist Economics 16*(2): 1-26.

Sonnenberg, B., M. Riediger, C. Wrzus, and G.G. Wagner (2012). Measuring Time Use in Surveys – Concordance of Survey and Experience Sampling Measures. *Social Science Research* 41(5): 1037-1052.

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.

Thompson, L. (1991). Family Work. Women's Sense of Fairness. *Journal of Family Issues* 12(2): 181-196.

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 Labor: 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 1: TABLES OF MAIN RESULTS WITH STANDARD ERRORS

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

| 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 | -1.106** | -1.194** | -1.012** | -0.190 |
| | (0.215) | (0.306) | (0.277) | (0.396) |
| Resp. does > 50% of housework | -1.539** | -1.699** | -1.387** | -0.306 |
| | (0.207) | (0.283) | (0.273) | (0.372) |
| Hours of paid work (ref: both full-time) | | | | |
| Both part-time | -0.167 | -0.188 | -0.094 | -0.096 |
| | (0.174) | (0.243) | (0.214) | (0.299) |
| Resp. full-time, partner part-time | -0.726** | -1.030** | -0.365 | -0.671* |
| | (0.164) | (0.232) | (0.203) | (0.299) |
| Resp. part-time, partner full-time | -0.287 | -0.355 | -0.179 | -0.166 |
| | (0.171) | (0.238) | (0.223) | (0.311) |
| Hourly pay (ref: about equal) | | | | |
| Resp. hourly pay half that of partner | -0.033 | 0.015 | -0.060 | 0.063 |
| | (0.129) | (0.186) | (0.169) | (0.247) |
| Resp. hourly pay double that of partner | -0.148 | -0.101 | -0.158 | 0.059 |
| | (0.134) | (0.183) | (0.182) | (0.248) |
| Paid help with housework (ref: none) | | | | |
| Cleaner one morning per week | -0.014 | -0.015 | -0.050 | 0.027 |
| | (0.108) | (0.156) | (0.136) | (0.201) |
| Presence of children (ref: none) | | | | |
| 6-month-old child | -0.005 | 0.157 | -0.238 | 0.394 |
| | (0.152) | (0.221) | (0.197) | (0.293) |
| 5-year-old child | 0.034 | 0.306 | -0.273 | 0.585* |
| | (0.156) | (0.218) | (0.196) | (0.279) |
| 15-year-old child | 0.322* | 0.315 | 0.307 | 0.001 |
| | (0.154) | (0.229) | (0.200) | (0.303) |
| Interactions: Resp. does < 50% housework X | | | | |
| Both part-time | 0.313 | 0.337 | 0.238 | 0.106 |
| | (0.202) | (0.280) | (0.263) | (0.365) |
| Resp. full-time, partner part-time | 1.249** | 1.537** | 0.921** | 0.633 |
| | (0.194) | (0.270) | (0.262) | (0.370) |
| Resp. part-time, partner full-time | 0.533** | 0.460 | 0.571* | -0.113 |
| | (0.197) | (0.273) | (0.269) | (0.370) |
| Resp. hourly pay half that of partner | 0.009 | -0.082 | 0.077 | -0.154 |
| | (0.158) | (0.228) | (0.201) | (0.297) |
| Resp. hourly pay double that of partner | 0.438** | 0.458* | 0.352 | 0.091 |
| | (0.161) | (0.224) | (0.222) | (0.309) |
| Cleaner one morning per week | 0.105 | 0.092 | 0.153 | -0.062 |
| | (0.134) | (0.186) | (0.175) | (0.243) |
| 6-month-old child | -0.034 | -0.057 | 0.066 | -0.097 |
| | (0.185) | (0.261) | (0.254) | (0.360) |
| 5-year-old child | -0.118 | -0.333 | 0.157 | -0.494 |
| | (0.185) | (0.260) | (0.235) | (0.334) |
| 15-year-old child | -0.578** | -0.519* | -0.601* | 0.100 |
| | (0.182) | (0.255) | (0.244) | (0.341) |

| | | | | |
|---|---|---|---|---|
| Interactions: Resp. does > 50% housework X | | | | |
| Both part-time | 0.088 | 0.018 | 0.156 | -0.131 |
| | (0.200) | (0.285) | (0.255) | (0.368) |
| Resp. full-time, partner part-time | 0.531** | 0.771** | 0.261 | 0.505 |
| | (0.194) | (0.278) | (0.244) | (0.360) |
| Resp. part-time, partner full-time | 1.062** | 1.313** | 0.760** | 0.529 |
| | (0.208) | (0.290) | (0.269) | (0.378) |
| Resp. hourly pay half that of partner | 0.167 | 0.049 | 0.299 | -0.244 |
| | (0.156) | (0.222) | (0.214) | (0.307) |
| Resp. hourly pay double that of partner | 0.084 | -0.030 | 0.180 | -0.217 |
| | (0.162) | (0.225) | (0.231) | (0.322) |
| Cleaner one morning per week | 0.352** | 0.453* | 0.256 | 0.198 |
| | (0.130) | (0.185) | (0.165) | (0.237) |
| 6-month-old child | -0.034 | -0.178 | 0.218 | -0.397 |
| | (0.187) | (0.265) | (0.247) | (0.358) |
| 5-year-old child | -0.079 | -0.370 | 0.282 | -0.650 |
| | (0.183) | (0.255) | (0.237) | (0.335) |
| 15-year-old child | -0.391* | -0.358 | -0.381 | 0.038 |
| | (0.187) | (0.266) | (0.256) | (0.365) |
| Age | -0.044** | -0.043** | -0.040* | -0.003 |
| | (0.012) | (0.016) | (0.017) | (0.023) |
| Age squared | 0.0004** | 0.0004** | 0.0004** | -0.00003 |
| | (0.0001) | (0.0001) | (0.0002) | (0.0002) |
| Observations | 4,547 | 2,476 | 2,071 | 4,547 |
| Number of individuals | 1,541 | 841 | 700 | 1,541 |
| Wald test $\chi^2$ | 425.59 | 312.34 | 158.11 | 479.93 |
| Wald *p*-value | .000 | .000 | .000 | .000 |
| Rho | 0.473 | 0.434 | 0.518 | 0.472 |

*Notes*: Asterisks denote significance levels. ** $p < .01$, * $p < .05$. Standard errors are robust to correlation between individuals belonging to the same household.
(3) Model with equal coefficients across gender and using the pooled sample of women and men,
(4) 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 faction of the variance of the unobserved component explained by the random effect.

**APPENDIX 2**

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

| 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 |
| | (0.184) | (0.250) | (0.248) | (0.335) |
| Resp. does > 50% of housework | -1.270** | -1.355** | -1.217** | -0.258 |
| | (0.175) | (0.232) | (0.243) | (0.313) |
| Hours of paid work (ref: both full-time) | | | | |
| Both part-time | -0.149 | -0.178 | -0.080 | -0.106 |
| | (0.150) | (0.204) | (0.197) | (0.263) |
| Resp. full-time, partner part-time | -0.621** | -0.843** | -0.348 | -0.554* |
| | (0.138) | (0.188) | (0.186) | (0.254) |
| Resp. part-time, partner full-time | -0.266 | -0.308 | -0.183 | -0.135 |
| | (0.146) | (0.199) | (0.201) | (0.269) |
| Hourly pay (ref: about equal) | | | | |
| Resp. hourly pay half that of partner | -0.035 | -0.003 | -0.057 | 0.041 |
| | (0.110) | (0.154) | (0.150) | (0.210) |
| Resp. hourly pay double that of partner | -0.136 | -0.089 | -0.151 | 0.053 |
| | (0.112) | (0.147) | (0.161) | (0.207) |
| Paid help with housework (ref: none) | | | | |
| Cleaner one morning per week | -0.005 | -0.040 | -0.005 | -0.047 |
| | (0.091) | (0.127) | (0.123) | (0.172) |
| Presence of children (ref: none) | | | | |
| 6-month-old child | -0.024 | 0.102 | -0.233 | 0.325 |
| | (0.126) | (0.174) | (0.176) | (0.245) |
| 5-year-old child | 0.013 | 0.247 | -0.269 | 0.517* |
| | (0.131) | (0.178) | (0.176) | (0.238) |
| 15-year-old child | 0.276* | 0.314 | 0.217 | 0.117 |
| | (0.132) | (0.191) | (0.181) | (0.262) |
| Interactions: Resp. does < 50% housework X | | | | |
| Both part-time | 0.262 | 0.300 | 0.190 | 0.131 |
| | (0.171) | (0.231) | (0.236) | (0.314) |
| Resp. full-time, partner part-time | 1.021** | 1.210** | 0.808** | 0.520 |
| | (0.161) | (0.215) | (0.233) | (0.309) |
| Resp. part-time, partner full-time | 0.456** | 0.383 | 0.507* | -0.084 |
| | (0.166) | (0.225) | (0.237) | (0.314) |
| Resp. hourly pay half that of partner | 0.019 | -0.043 | 0.069 | -0.110 |
| | (0.133) | (0.185) | (0.176) | (0.249) |
| Resp. hourly pay double that of partner | 0.367** | 0.366* | 0.313 | 0.070 |
| | (0.134) | (0.179) | (0.194) | (0.256) |
| Cleaner one morning per week | 0.068 | 0.084 | 0.089 | 0.004 |
| | (0.111) | (0.149) | (0.156) | (0.204) |
| 6-month-old child | 0.011 | -0.015 | 0.120 | -0.104 |
| | (0.152) | (0.206) | (0.223) | (0.298) |
| 5-year-old child | -0.082 | -0.279 | 0.184 | -0.468 |
| | (0.155) | (0.208) | (0.212) | (0.282) |
| 15-year-old child | -0.478** | -0.461* | -0.470* | -0.021 |
| | (0.153) | (0.208) | (0.217) | (0.290) |

| | | | | |
|---|---|---|---|---|
| Interactions: Resp. does > 50% housework X | | | | |
| Both part-time | 0.094 | 0.040 | 0.158 | -0.112 |
| | (0.170) | (0.234) | (0.229) | (0.313) |
| Resp. full-time, partner part-time | 0.457** | 0.638** | 0.246 | 0.420 |
| | (0.162) | (0.222) | (0.215) | (0.298) |
| Resp. part-time, partner full-time | 0.860** | 1.013** | 0.658** | 0.406 |
| | (0.172) | (0.234) | (0.236) | (0.317) |
| Resp. hourly pay half that of partner | 0.135 | 0.054 | 0.243 | -0.166 |
| | (0.130) | (0.179) | (0.185) | (0.253) |
| Resp. hourly pay double that of partner | 0.094 | 0.013 | 0.166 | -0.153 |
| | (0.135) | (0.180) | (0.202) | (0.265) |
| Cleaner one morning per week | 0.268* | 0.363* | 0.181 | 0.216 |
| | (0.107) | (0.148) | (0.145) | (0.199) |
| 6-month-old child | -0.002 | -0.116 | 0.229 | -0.338 |
| | (0.153) | (0.209) | (0.214) | (0.295) |
| 5-year-old child | -0.061 | -0.310 | 0.263 | -0.565* |
| | (0.153) | (0.206) | (0.209) | (0.280) |
| 15-year-old child | -0.346* | -0.382 | -0.275 | -0.125 |
| | (0.158) | (0.219) | (0.227) | (0.311) |
| Age | -0.037** | -0.035** | -0.036* | -0.002 |
| | (0.010) | (0.013) | (0.016) | (0.020) |
| Age squared | 0.000** | 0.000** | 0.000** | -0.000 |
| | (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 $\chi^2$ | 397.52 | 299.79 | 158.49 | 447.02 |
| Wald *p*-value | .000 | .000 | .000 | .000 |
| Rho | 0.523 | 0.498 | 0.553 | 0.524 |

*Notes*: Asterisks denote significance levels. ** $p < .01$, * $p < .05$. Standard errors are robust to correlation between individuals belonging to the same household.
(5) Model with equal coefficients across gender and using the pooled sample of women and men,
(6) 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 faction of the variance of the unobserved component explained by the random effect.