UMR 8184
Thema
THéorie Économique, Modélisation et Applications

THEMA Working Paper n°2017-10
Université de Cergy-Pontoise, France

Efficiency versus Gender Roles and Stereotypes: An Experiment in Domestic Production

Hélène Couprie, Elisabeth Cudeville, Catherine Sofer

April 2017

UNIVERSITÉ de Cergy-Pontoise
CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Efficiency versus Gender Roles and Stereotypes: An Experiment in Domestic Production*

Hélène Couprie † Elisabeth Cudeville‡ Catherine Sofer§

Abstract

Empirical studies cast doubts upon the efficiency assumption made in standard economic models of household behavior. The allocation of time among men and women between market and household work is highly differentiated by gender. In this paper we examine whether couples deviate from efficiency in household production, using an experimental design. We compare the allocation of gendered vs. neutral tasks. Our results show that women in the household overspecialize in "feminine tasks" and men in "masculine tasks" compared to what their comparative advantage would require, hence revealing the influence of gender roles and stereotypes on the couples' behavior.

Keywords: Stereotypes, gender roles, household models, gender inequalities.

JEL Classification: D13 - J16 - J22- C91- C92

*We thank P. Apps, P-A Chiappori, the participants of the June 2014 Workshop "Economics of Gender" at Nice and the anonymous referees for very helpful comments and suggestions. We also thank all the experimenters who helped us in running this experiment. Financial support from the French National Research Agency (ANR "Ginhdila") is gratefully acknowledged. All errors remain ours.

†THEMA, University of Cergy-Pontoise.

‡UMR Development & Societies, IEDES-University of Paris 1 Panthéon-Sorbonne.

§PSE and University of Paris 1 Panthéon-Sorbonne.

1 Introduction

In most countries, the allocation of time-uses remains highly differentiated by gender (Goldschmidt-Clermont and Pagnossin-Aligisakis 1995). Gender inequalities in the sharing of domestic work remain persistent despite the decline in the gender pay gap (Aguiar and Hurst 2007). This observation is puzzling under standard economic assumptions, and is far from being satisfactorily explained by economists. The aim of this paper is to investigate productive efficiency in the household through an incentivized experiment involving established couples and further explore the role of task preferences, gender roles and stereotypes in spouses' specialization choices.

In Europe, men spend 55 to 65 percent of their working time in market work, and 35 to 45 percent in household work (Aliaga and Winqvist 2003). Women spend 60 to 70 percent of their working time on domestic tasks and 30 to 40 percent in paid-work and they still carry most of the burdens of household work[1]. This biased specialization of women in household duties and of men in market activities is very likely to interact strongly with women's situation in the labor market. Persisting gender differences in wage rates (Blau and Kahn 2006) and earned incomes, through shorter work duration, as well as through flatter careers, could be the consequence of the unequal sharing of domestic work within the family.

Productivity differences between men and women in market activities could explain why their time-allocation decisions differ. However, empirical studies show that variables reflecting productivity in the labor market, such as wage rates and education, are far from being the only factors driving the gendered division of labor (Hersch and Stratton 1994; Aronsson et al. 2001; Anxo et al. 2002; Rapoport and Sofer 2005; Kalenkoski et al. 2009). Using the French Time-Use Survey, Rizavi and Sofer (2010) show that when a woman earns a higher wage than her partner, there is no role reversal. Sevilla-Sanz et al. (2010) find similar results in Spanish data. In the US and in Australia, the gender gap in non-market work is even

[1]In France, in 2009, women contributed approximately 80 percent of domestic chores (Régnier-Loilier 2009).

higher when the woman earns more than her partner (Bertrand et al. 2013; Bittman et al. 2003). These observations are hard to reconcile with standard household models that assume Pareto-efficiency in intra-household decisions. Furthermore, a number of studies question and reject the validity of the efficiency assumption, especially on the household production side (Udry 1996; Duflo and Udry 2004).

Another way to reconcile the empirical evidence described above and the efficiency assumption is to assume some gender differences in marginal productivity of labor in home activities. Women could be more productive than men at home, and their productivity advantage could more than compensate for the wage gap when they are better paid than their partner. This is the assumption made by Becker (1981), based – possibly besides biological abilities– on the different training received by girls and boys. However today, in developed countries at least, though toys and games are still highly gendered, it is difficult to argue that specific skills are needed to perform household chores given the widespread use of market substitutes for domestic goods and household appliances. Moreover, most fathers are now involved in the caring of children at all ages, including newborns. Therefore, excluding the very limited time period of pregnancy, and possibly breastfeeding, systematic gender productivity differences in domestic tasks and the raising of children do not seem likely.

Beside productivity differences, persistent gender differences in preferences could explain the aforementioned puzzle. Women, on average, may "like" to perform domestic tasks more than men, due to habit, education or socialization. A consumption component would then appear on the production side of household decision-making. Non-monetary costs (or advantages) are then added (or subtracted) to monetary costs in the household profit function. To be consistent with the observed division of labor, the difference between men and women in these non-monetary costs should be high enough to compensate for any difference in wages, when women's wages are higher.

Social norms could also be an explanation. Rules of behavior dictated by society such as "gender roles" could impose additional gains or costs to direct preferences when the indi-

vidual does or does not comply with these rules (Akerlof and Kranton 2000).[2] In this case, households might choose rationally to deviate from first-best efficiency. Duflo and Udry (2004) reject the efficiency assumption in the context of Côte d'Ivoire and show that the intra-household distribution of income and its uses are strongly constrained by social norms.

Inefficiencies in domestic production could also stem from gender stereotypes. Gender stereotypes are the sets of beliefs grounded in society that link personal attributes to the social categories "women" and "men". For example, people may believe that women are more talented than men in domestic tasks and child care.[3]. Until now, empirically, the effects of stereotypes on individual behavior have mainly been explored in the field of experimental psychology.[4]

In this paper, we propose an experimental approach to test the efficiency of the division of labor within the family, and investigate the role of preferences, gender roles and gender stereotypes in explaining the observed deviations from productive efficiency. Indeed, in practice, since household production is generally not observed in datasets[5], convincing micro-econometric tests of efficiency cannot be implemented. Although time inputs from both partners can be recovered through time-use data, neither non-market inputs nor outputs can be observed. Thus, domestic production functions can never be estimated in a satisfactory way, and the different explanations of the paradox stressed above cannot be tested. The experimental approach overcomes these difficulties by allowing a direct measure of individual domestic productivities and the control of the production technology.

In the experiment, established couples were asked to complete day-to-day tasks within a maximum allotted time, and the free time saved by each spouse over the allotted time was

[2]Household models including social norms have been developed by Lundberg and Pollak (1993), Carter and Katz (1997) and Cudeville and Recoules (2015).

[3]Coate and Loury (1993) and Sofer and Thibout (2011) proposed theoretical models of stereotypes to explain persistent inequalities in the wages of minority workers for the former and in the sharing of household tasks for the latter

[4]Note, however, Coffman (2014) who ran a lab experiment that shows the effect of gender stereotypes on individual's decisions to contribute her ideas to a group.

[5]Except for some agricultural households in developing countries.

paid to the couple, possibly at different rates.[6] Thus, couples had an incentive to minimize the time spent on tasks as it represented a fixed cost that reduced their paid-time and hence their monetary payoff. Individual productivities on tasks were measured in separate sessions to calibrate the optimal allocation. It was then possible to verify whether spouses allocated their time efficiently across tasks according to their comparative advantages when working together (that is, maximized their joint monetary payoff). In order to investigate the role of gender norms and stereotypes in deviations from efficiency, the experiment contained a 2×2 design: half of the couples faced two neutral tasks (stuffing envelopes, completing forms) and the other half two highly gender stereotyped tasks (folding socks, screwing on metal brackets), while in each condition, half of the couples were informed of their individual productivities and the other half were not. Information about preferences and beliefs was gathered through individual questionnaires during the course of the experiment.

A vast body of literature has now analyzed gender differences using an experimental framework (see the survey by Croson and Gneezy 2009). Besides, a number of experiments involved established couples for getting inside the household and exploring various aspects of the household decision-making, such as efficiency, bargaining, risk attitude, time preference (see Munro 2015 for a comprehensive survey). The present paper belongs to this strand of literature but is new on three grounds. First, it explores an aspect of efficiency that, to our knowledge, has not been studied experimentally to date: the production side of the household decision-making and the division of labor. Second, unlike most experimental studies on household efficiency which rely on experimental games, such as trust, voluntary contribution or modified dictator games (e.g. Peters et al. 2004; Chao and Kohler 2007; Munro et al. 2010, 2014; Beblo et al. 2015; Cochard et al. 2016), subjects had to work on concrete day-to-day tasks. Rather than featuring "intellectual" real effort tasks, such as solving mazes or mathematical problems, we implemented manual work tasks in order to mimic, as much as possible, real household production.[7] Lastly, while in a typical experiment couple's members

[6] We thus allow for differences in opportunity cost of time between male and female partners.

[7] Few papers in experimental economics used real effort working tasks, but none in the context of the

are asked to make decisions together, to share information, or to make decisions separately, the design proposed here does not force a couple to adopt a particular way of making their decisions. The latter two points are methodological in nature but important in terms of external validity, as pointed by Munro (2015).

Our main results are the following. Although comparative advantages partially drive spouses' specialization choices, on average, couples fail to attained full maximization of their money payoff. Moreover, inefficiencies are exacerbated in the direction dictated by gender roles and stereotypes when they face gendered tasks. The information provision to the couple of individual productivities on tasks improves efficiency in the gendered condition, which confirms the role of stereotypical believes in task sharing within the couple, although differences in preferences, controlled for using data collected from individual questionnaires, have no significant effect.

The remainder of the paper is organized as follows. Section 2 describes the conceptual framework that guides the experimental design. Section 3 details the experimental design: the setting and treatments. Section 4 presents the experimental results and the empirical analysis of these results. Section 5 concludes.

2 Conceptual Framework: Designing the Experiment

2.1 The Theoretical Model

Let us consider a couple made up of a man (m) and a woman (f). Each partner is endowed with T units of time. The level of domestic good production is fixed, so time inputs in domestic tasks are modeled as implying opportunity costs that need to be paid and shared within the couple. The opportunity cost of domestic production may differ between spouses

study of couple behavior. In Fahr and Irlenbusch (2000), e.g., subjects had to crack walnuts and in Falk and Ichino (2006), they were asked to fill letters in envelopes.

depending on the female and male wage rates prevailing on the labor market.[8] By sharing domestic tasks according to Ricardian comparative advantages, the couple can minimize the opportunity cost of domestic production and hence maximizes its potential income.

Formally, we assume that the partners have to perform two tasks consisting in producing a fixed quantity $\bar{q}_i$ of two goods $i = 1, 2$. The production function of each good is assumed linear with respect to the time inputs of each spouse, which are considered substitutable. The time spent by member $j = m, f$ in task i is denoted by t_i^j. The marginal productivity of labor in each task may differ between spouses but is assumed not to depend on time spent and quantity produced. Hence, partner j's productivity in task i, denoted by α_i^j, is fixed. The production processes of the two tasks are assumed independent. Completing the two tasks within the allotted time, whoever performs them, is the condition for partners to enter and value their time on the labor market, where the female and male wages are respectively set at w^f and w^m.

An efficient couple would agree on an allocation of time across tasks that maximizes their potential income (Chiappori 1997, Apps and Rees 1997). Under the standard assumption that the production decisions of the couple are separable from its consumption decisions, the couple's problem can be written as follows:

$$\begin{cases} \max w^f(T - t_1^f - t_2^f) + w^m(T - t_1^m - t_2^m) \\ \text{s.t.} \quad \alpha_1^f t_1^f + \alpha_1^m t_1^m \geq \bar{q}_1 \\ \qquad \alpha_2^f t_2^f + \alpha_2^m t_2^m \geq \bar{q}_2 \\ \qquad t_1^f, t_2^f, t_1^m, t_2^m \geq 0 \end{cases} \tag{1}$$

We restrict our attention to cases where each spouse has enough time to complete alone

[8] If time spent at home and time spent in paid work can be freely substituted for one another, the market wage rate measures the opportunity cost of an hour at home.

both domestic tasks in the allotted time, satisfying the following conditions:[9]

$$\frac{\bar{q}_1}{\alpha_1^f} + \frac{\bar{q}_2}{\alpha_2^f} < T \quad \text{and} \quad \frac{\bar{q}_1}{\alpha_1^m} + \frac{\bar{q}_2}{\alpha_2^m} < T \tag{2}$$

The first-order conditions for program (1), reported in Appendix A, define three specialization patterns according to the comparative advantage of spouses, which in turn depends on their relative productivity in tasks and wages (see Table A.1 in Appendix A for details):

i) If the woman's relative wage rate is lower than her relative productivity in both tasks, then the couple's efficient allocation is for the woman to complete both tasks alone, while the man does not participate in domestic production and saves all his time endowment. We refer to this pattern as the "Woman Does Everything" regime, denoted by WDE.

ii) If the woman's relative wage rate is higher than her relative productivity in both tasks, then the optimal sharing is for the man to perform the two tasks alone, and for the woman to save all her available time. We refer to this pattern as the "Man Does Everything" regime, denoted by MDE.

iii) If the woman's relative wage rate lies between her relative productivities in both tasks, the efficient allocation is that each partner takes charge of the task for which he/she has a comparative advantage. We refer to this pattern as the "Man and Woman Share Tasks" regime, denoted by MWS.

2.2 Measuring Productive Inefficiency

The couple will be considered efficient if it behaves in accordance with the above model predictions. The total output being fixed for each task, the Pareto-efficiency of task allocation within the couple can be assessed using the output produced by one of the two partners only.

[9]The experimental design was properly calibrated to meet these conditions.

For each task, the male output in any couple work session is simply the total required output minus the female output. Formally, let us define the vector $\mathbf{q} = (q_1, q_2)$, where $q_i \equiv q_i^f / \bar{q}_i$ denotes the quantity of good i produced by the woman, normalized by the total output required for this task. Let us denote by $\mathbf{q}^* = (q_1^*, q_2^*)$ the efficient output vector. It depends on the structure of comparative advantage which in turn depends on partners' wages and productivities in tasks: $\mathbf{q}^*$ is equal to (1,1) in the WDE regime; (0,0) in the MDE regime; (1,0) in the MWS regime if the woman has a comparative advantage in task 1 and $(0, 1)$ if she has a comparative advantage in task 2.

To characterize productive inefficiencies we define the indicator function d_1, that takes value 1 if the couple decision is Pareto-efficient and 0 otherwise:

$$d_1 = \begin{cases} 1 & \text{if } \mathbf{q} = \mathbf{q}^* \\ 0 & \text{if } \mathbf{q} \neq \mathbf{q}^* \end{cases} \tag{3}$$

By summing the d_1 indicator over couples, we get the proportion of efficient couples, which we denote by $\bar{d_1}$.

To capture the degree of productive inefficiency within the couple, we use the Manhattan distance d_2, between the actual and efficient allocations:[10]

$$d_2 = |q_1 - q_1^*| + |q_2 - q_2^*| \tag{4}$$

Furthermore, in the gendered condition, we will orient this measure in order to identify whether the deviation from efficiency occurs in the direction expected from the gender stereotype: the woman does more socks (task 1) and/or less metal brackets (task 2) than what efficiency requires. In this case, the d_2 distance is denoted by d_2^{exp}, and in the other case,

[10]Other measures of distance as the Euclidian distance give similar results.

where the deviation goes in the opposite direction, the distance is denoted by d_2^{unexp}:

$$d_2 = \begin{cases} d_2^{exp} & \text{if } q_1 \geq q_1^* / \text{or } q_2 \leq q_2^* \\ d_2^{unexp} & \text{otherwise} \end{cases} \tag{5}$$

Averaging these three measures d_2, d_2^{exp} and d_2^{unexp} over couples will allow us to isolate the effect of gender norms and stereotypes in explaining couples' deviations from efficiency, since we get: $\bar{d}_2 = \bar{d}_2^{exp} + \bar{d}_2^{unexp}$. Hence $\bar{d}_2^{exp}$ gives the share of the average deviation of couples' choices from efficiency that can be explained by gender roles and/or stereotypes.

3 Experimental Design

The experimental design described below was intended to mimic the sharing of highly-gender stereotyped household tasks and compare it with the sharing of neutral tasks to assess the role of gender roles and stereotypes in household productive inefficiencies. Particular attention was paid to the information setting of the couples and individual preferences for tasks.

3.1 General Setting

We ran 24 experimental sessions in two locations: 12 at the University of Paris 1 Panthéon-Sorbonne, and 12 at the University of Cergy-Pontoise (a medium-sized city in the Paris suburbs) from March 2012 to January 2013. The sample comprised 81 established couples[11] recruited through a public announcement in the media (newspapers, radio and television) inviting participation in an economic experiment for money. Volunteers were informed that their couple could earn between €50 and €100 for a session of approximately 2 hours, but had no information about the purpose or the content of the experiment. To participate,

[11] In our case, using established couples was necessary since we were interested in observing the impact of social gender norms and stereotypes in the specific context of the family which is presumably characterized by a high level of communication and cooperation. If gender bias and stereotypes come into play in the intimacy of the family unit, we may expect them to affect any kind of relationships in which information sharing is weaker.

partners had to have cohabited for a continuous period of at least one year.[12] Registration for scheduled experimental sessions took place online through a web site[13] and ensured that participants registered only once. Sessions were held on Saturdays, weekday afternoons or lunchtimes, to reduce the cost of participation and attract people with various socio-demographic backgrounds.

Each experimental session lasted approximately 2 hours, involved a maximum of 4 couples, and was conducted by 1 or 2 of the authors assisted by 4 to 5 experimenters. Participants had to work individually and as a couple on two manual tasks. To complete this design, subjects were asked to fill out three individual questionnaires during the course of the experiment: the first – Q1 – to elicit their beliefs about their own abilities to perform the required tasks as well as those of their partner, the second – Q2 – concerning their socio-economic characteristics, and the third – Q3 – to assess their preferences on the accomplished tasks (questionnaires Q1 and Q3 are reported in Appendix B).

The time performances in the accomplishment of tasks were measured by experimenters who were also in charge of reporting production and ensuring that the instructions were perfectly understood and followed. The instructions[14] were given orally and simultaneously projected on a screen. This experiment involved substantial gains and required time and sustained work effort from participants, who appeared to take it very seriously. In this respect, it produces quality data. A problematic issue related to the experimental design stems from the possible bias introduced by the presence of experimenters (1 for 2 participants). Participants had to make their decisions freely without feeling any pressure from observation. The team of experimenters was aware of this issue and trained to remain discrete and keep interactions with the subjects as minimal as possible.[15]

[12]This condition was checked through questionnaire Q2. The couple status was also checked using specific questions regarding family in-law

[13]The web site www.expeco.fr was maintained by the Parisian Experimental Economics Laboratory (LEEP). A phone number was available to provide assistance if needed.

[14]Instructions and questionnaire Q2 are available from the authors.

[15]We further checked whether the gender of experimenters had an impact upon the results (see section 4.2, Table 8).

Table 1: Summary Statistics of the Sample

(a) Individual Characteristics

Variables*	**Full Sample**	**Females**	**Males**
	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	40.6 (12.2)	39.6 (12.1)	41.5 (12.2)
Active (percent)	85.8	80.2	91.4
Personal Bank Account (percent)	80.6	84.0	77.2
Diploma Level (percent)			
Lower secondary	24.1	18.4	29.6
Upper secondary	20.4	19.8	21.0
Short-cycle tertiary	14.2	16.0	12.3
Bachelor's or equivalent	16.7	19.8	13.6
Master's or higher level	24.7	25.9	23.5
Observations	162	81	81

(b) Couples' Characteristics

Variables*	**Full Sample**	**Cergy**	**Paris**
	Mean (SD)	Mean (SD)	Mean (SD)
Average spouses' age (years)	40.6 (12.0)	40.3 (12.4)	40.9 (11.6)
Couple duration (years)	12.5 (11.3)	11.9 (11.6)	13.2 (11.1)
Childless (percent)	50.6	47.7	54.1
Married (percent)	45.7	45.5	45.9
Civil union ("Pacs") (percent)	17.3	13.6	21.6
Neither married nor "Pacsed" (percent)	37.0	40.9	32.4
Joint Account (percent)	64.8	63.6	66.2
Observations	81	44	37

Notes: *All variables were collected separately for male and female partners.

Since participation was voluntary and the characteristics of couples who did not participate were not observed, a participation bias in the recruitment process cannot be excluded. Still, the sample displays substantial heterogeneity. Table 1 provides summary statistics of the sample from the individual socioeconomic questionnaire Q2 completed in the course of the experiment. The average age of partners ranged from 27 to 70.5 years, with a mean of 40.6 years. In average, women were younger than their partner with an average age gap equal to 2 years, which is standard in French population data.

The couples were stable: they have been living together for 12 years on average. They had average employment rates and education levels similar to those of the general population of couples in the Ile-de-France region. Selection according to these observable characteristics was comparable in Cergy and in Paris. Consistently with the population data when comparing Paris to its suburb, in the Paris sample, couples were less often married, more often childless[16] and more often employed. However, our sample satisfactorily reflects the diversity of French couples. Among participating couples, only 2 reported having already participated in economic laboratory experiments. The two main participation motives declared by subjects were curiosity and money, for both men and women.

3.2 Treatments and Predictions

In order to explore and disentangle the effects of gender roles and stereotypes on the efficiency of task sharing within couples, we randomly assigned the 81 recruited couples to one of four treatment conditions in a 2×2 between-subjects design (see Table 2). Half the couples faced gendered tasks (gendered condition, G) and the other half – the control group – faced neutral tasks (neutral condition, N). Moreover, in each group, half of the couples received information about the performance of each partner on tasks (informed condition, I) and half did not (not-informed condition, NI).

Table 2: Experimental Design 2x2 (Nb. of Couples)

Conditions	Not informed (NI)	Informed (I)	All
Neutral (N)	21	20	41
Gendered (G)	21	19	40
All	42	39	81

The first treatment (G/N) regarding the nature of the tasks was intended to reveal the potential effects of gender norms and stereotypes on the allocation of tasks between spouses.[17]

[16]Though the percentage of childless couples in both Cergy and Paris is very close to average, less couples in our sample were married compared to the average population of couples in the Ile-de-France region.

[17]The treatment was chosen to be applied between subjects. A within treatment has been excluded as

Gendered tasks are viewed as being stereotypically attributed to a specific gender, whereas neutral tasks do not have this property. In the gendered condition, the "feminine task" consisted of sorting through and matching up socks, and the "masculine task" of assembling metal brackets with bolts and screws. Our choice of an activity involving clothes for "the feminine task" and of a repair activity for the "masculine task" was based on European time-use surveys, which show that domestic tasks involving clothes as well as construction and repair activities are highly gendered, the former being performed mostly (and sometimes exclusively) by women, and the latter by men.[18] For the neutral tasks, we chose classical experimental tasks: a first task consisting in filling envelopes and a second task consisting of completing nominative forms with the relevant phone number.[19] As we expected that gender stereotypes and norms could influence the task allocation decisions of couples in a way that possibly contradicts the logic of comparative advantage, we made the following predictions:

Prediction 1. *Couples in the gendered condition compared to those in the neutral-control condition will deviate more often and/or to a larger extent from efficiency in their task sharing.*

Prediction 2. *In the gendered condition, men will overspecialize in the "masculine task" and women in the "feminine task" compared to what their comparative advantages dictate.*

However, the same observations could be made by the sole effect of gendered preferences: if women actually preferred "the feminine task" and men "the masculine task", results in line with predictions 1 and 2 would be obtained. Implicitly, standard efficient models of the household – which rely on the separability assumption of household production from consumption decisions (**see** Chiappori 1997) – ignore the role of individual preferences for

too time-consuming, since it would have forced couples to perform four different tasks.

[18] In Europe, for example, Winqvist (2004) compared 10 countries and showed that the share of women in the total household time devoted to laundry ranged from 80 percent in Sweden to 100 percent in Slovenia. Conversely, "construction and repairs" is a masculine activity, as, on average, the sharing of time is around 85 percent for men and 15 percent for women in Europe.

[19] Indeed, "household management tasks" are among the tasks the most equally-shared between genders. Among French couples (Time-Use Survey 2009-2010), household management is carried out by men (resp. women) for 54 (resp. 46) percent of couples.

tasks. Hence, to assess the role of gender roles and stereotypes on productive efficiency within the couple, we have to control for individual preferences using the related information collected by questionnaires. If the standard separability assumption was to be confirmed, then we would expect that:

Prediction 3. *Preferences of subjects over tasks would not play a major role in task allocation within couples.*

The second treatment (I/NI) regarding information about partners' performances on tasks was implemented to isolate the potential effect of gender stereotypes on the division of tasks between spouses from that of gender roles. For half of the sample – the informed group – the individual performances of each spouse for both tasks were indicated to the couple at the beginning of joint work sessions, although for the other half – the non-informed group – this information was not given and could thus remained private. If deviations from efficiency came from false beliefs based on gender stereotypes, providing the true information to both partners regarding their productivity should lead them to revise their beliefs, at least partially, so as to decrease inefficiencies. Conversely, if the division of tasks between spouses was explained only by gender roles, there would be a cost in deviating from the social standards, whatever the true performance in completing tasks. In this case, what is important for subjects is "doing gender" (West and Zimmerman 1987). If people consider, to some extent, that it is the "role" of women to deal with certain tasks, while it is that of men to deal with others, they will choose to share tasks in this way, even if this implies a monetary loss. Information plays no role then. Hence, we made a fourth prediction:

Prediction 4. *If gender roles rather than stereotypes (mistaken beliefs about gender-specific productivity for gendered tasks) are what drive the over-allocation of the female time in the "feminine task" and of the male time in "the masculine task", and hence inefficiencies, then providing information should not significantly impact the couples' behavior.*

3.3 Tasks and Production Rules

The tasks to be performed were chosen according to the following requirements. They should be simple enough not to leave room for variations in quality. Performances of participants should be different but rather stable in time, so we could rule out learning or tiredness effects. Moreover, the tasks had to be manageable manually in a simple room, and not too costly. Finally, the production had to be easy to count and divisible.

3.3.1 Gendered Tasks

The female-oriented gendered task consisted in sorting out and rolling up 11 pairs of matched socks from a basket containing 11 potential pairs and 10 unmatched socks. The socks, mixed up in the basket, were of different colors, sizes and patterns (the material used for the task is shown in Appendix C, Fig. C.1a). The basket was changed between each work session.[20] The production level to be reached was 11 pairs of socks, within a maximum time of 3 minutes and 30 seconds.

The male-oriented gendered task consisted in assembling 2 metal brackets with 2 nuts and 2 screw bolts to form a 'T' (see Appendix C, Fig. C.1b). The nuts should be tightened by hand without any play in order to obtain a stable T. The material was presented in a flat rectangular container with more nuts and screw bolts than necessary, but the exact number of metal brackets. The production level to be reached was 5 T shapes, within a maximum time of 3 minutes and 30 seconds.

3.3.2 Neutral Tasks

To perform the first neutral task, participants had to fill 5 envelopes, each with two sheets of paper of different colors, appropriately folded in three, and seal them. The material put on the table contained exactly the right number of envelopes and two piles of paper of different

[20]The participants were not allowed to empty the basket on the table to reconstitute the pairs of socks, they had to work in the basket and could not get the matched socks out of the basket before having folded them into a ball.

colors (see Appendix C.2a). The required production level of 5 envelopes had to be reached in a maximum time of 3 minutes and 30 seconds in individual sessions.

The second neutral task was to complete nominative forms with the appropriate phone numbers which had to be found in a list of 50 names, not in alphabetical order, matched with phone numbers. The material for this task consisted of pens, 5 nominative forms and a list of two pages of 50 names and phone numbers (see Appendix C.2b). The forms and the name list were different in each round. The required production level was to fill out the 5 nominatives forms with the appropriate phone numbers, in a maximum time of 3 minutes and 30 seconds in individual sessions.

3.3.3 Joint Production Rules for Couples

During couples' work sessions the required level of joint production for a couple was *i)* in the gendered setting, to assemble 10 metal brackets to form 5 'T' and roll up 11 pairs of matched socks in a maximum time of 7 minutes and *ii)* in the neutral setting, to produce 5 envelopes and fill out 5 nominative forms with the appropriate phone numbers in the same limited time of 7 minutes. The rules of joint production for a couple were precisely described to the participants with illustrative examples. In order to avoid complementarity between spouses' working times, the participants were not allowed to help each other in the production process of one good (the partners could not work together to produce one pair of matched socks, one 'T', one envelope or one nominative form). Thus, although allowed to work in parallel on the same task, partners had to work independently.

3.4 Detailed Implementation of the Experiment

Each experimental session proceeded in five stages, as described in Table 3.

In stage 0, the participants, gathered in the same room, were informed about the timeline of the experiment, the tasks to carry out and the general payment rules. Subjects were assured that confidentiality was guaranteed as well as anonymity through personal ID badges

Table 3: Timeline of an Experimental Session

0	**ALL PARTICIPANTS**		*In the same room.*
	Welcome and general instructions		
	GENDER/NEUTRAL TREATMENT randomized across sessions.		
1	**INDIVIDUAL WORK SESSION**		*Men and women separately in 2 rooms.*
	Training time	(10 mn)	
	Tasks 1 & 2		
	Measurement of individual productivities		
	Task 1	(3:30 mn)	Order of tasks randomized across participants.
	Task 2	(3:30 mn)	
	Questionnaire Q1		
	Questionnaire Q2		Randomized across stage 1 & 3.
2	**COUPLES' WORK SESSIONS**		*2 couples per room separated by screens.*
	INFORMED/NOT INFORMED TREATMENT, randomized across sessions.		
	Each couple receives a paper indicating:		
	– The 3 wage schemes (WDE, MDE, MWS)		Specific to the couple, in random order.
	– Individual performances at stage 1		For Informed sessions only.
	Round 1		
	Free discussion time	(5 mn)	
	Joint work on tasks 1 & 2	(7 mn)	
	Round 2		
	Free discussion time	(5 mn)	Wage schemes WDE, MDE, MWS in random order between rounds.
	Joint work on tasks 1 & 2	(7 mn)	
	Round 3		
	Free discussion time	(5 mn)	
	Joint work on tasks 1 & 2	(7 mn)	
3	**INDIVIDUAL WORK SESSION**		*Men and women separately in 2 rooms.*
	Measurement of individual productivities		
	Task 1	(3:30 mn)	
	Task 2	(3:30 mn)	Order randomized across participants.
	Questionnaire Q2		Randomized across stages 1 & 3.
	Questionnaire Q3		
4	**PAYMENT**		*Granted privately to each couple.*

randomly allocated on their arrival. They were also informed about the exchange rate of the experimental currency unit, the ECU, set at ECU35 per euro. They were explained that the payment was given jointly to the couple, in the presence of both partners, at the end of the experimental session.[21] At the recruitment stage, it was stated that participant couples would make a guaranteed €20, and from €0 to a feasible maximum of around €80 based purely on performance. We clearly communicated to the subjects that the payment would depend heavily on performance and that the sooner they could accomplish the tasks at each stage of the experiment, the more money the couple would earn. The experiment had 5 work sessions on the same two tasks (gendered or neutral, depending on the treatment condition): a first individual work session in stage 1, followed by three couples' work sessions in stage 2, and a last individual work session in stage 3, with short breaks in between. The two tasks to be accomplished were then presented in detail, on the basis of a demonstration performed with the same material as that used by the participants.

Stages 1 and 3 were intended to measure individual productivities on tasks. During these individual work sessions, men and women were separated in two different rooms. Subjects were randomly assigned to cubicles, spread out across the rooms and separated by screens.[22] The subjects were asked to accomplish each of the two required tasks within the allotted time, and were explicitly reminded that the faster they would work, the higher the gain of their couple would be. They first had a training period of 10 minutes to get familiar with the tasks. This training was intended to ensure a certain stability of individual performances, an assumption made in the theoretical model. The second productivity measure performed at stage 3 was precisely intended to check the stability of the individual productivity parameters. Participants had a maximum time of 3 minutes and 30 seconds to accomplish each task sequentially. This time and the required output for each task were calibrated in prior

[21] Spouses were unable to identify which share of the money payoff earned was attributable to their individual output or work effort, which greatly limited the issue of "ex post undoing" stressed in Munro 2015 and that corresponds to the fact that the spouse who contributed most to generating earnings might have a greater say about how to spend the money.

[22] A cubicle was made of three tables configured in U-shape with space for two subjects working back to back. This configuration allowed one experimenter to measure in parallel the performance of two subjects.

pilot sessions to ensure that any participant could complete the tasks within the allotted time.[23] An experimenter gave the starting signal to each participant by activating a desktop stopwatch, and recorded the time once the required output was reached. The stopwatch was placed in front of each subject in such a way that the subjects, but the subjects only (apart from the experimenter) could see their performances. Such discrete reporting was made to prevent competition between participants. The difference between the allotted time and the time actually spent by the subjects to complete the task was paid at an announced rate of ECU10 per second. The questionnaire Q1 aiming at eliciting beliefs was filled out by subjects at the end of stage 1, and the questionnaire Q3 concerning preferences for tasks at the end of stage 3.

In stage 2, the members of each couple were reunited to work on the same two tasks, but this time together. The couple had to complete the two tasks in a limited total time of 7 minutes. This treatment was repeated three times for each couple, with different wage schemes corresponding to the different specialization patterns MDE, WDE and MWS described in the theoretical section 2.1, in a random order. Using individual productivity measures obtained in stage 1, three vectors of wages (w^m, w^f) – setting the opportunity cost of time of each partner – were computed[24] to ensure that, over the three couples' work sessions, the efficient solution fell within each of the regimes (conditions are given in Appendix Table A.1).

Stage 2 was conducted as follows. The couples were first gathered in the same room to receive the instructions (tasks to be performed, rules of the game and payment scheme) and then spread out across the two rooms where they were randomly assigned to cubicles, isolated by screens, at the opposite corners of each room. The material needed to perform the tasks was displayed on the central table of each cubicle, and the partners were free to sit wherever they wanted inside this space. Before starting to work, the couples received a sheet of paper indicating the three wage vectors that would apply successively to the three

[23] No subject in the sample failed in accomplishing the tasks in the time limit.

[24] The wage vectors were computed automatically using a pre-programmed Excel-file at the end of stage 1, once the individual productivity measures had been collected from experimenters.

joint work sessions. In half of the experimental sessions (informed condition), the sheet also indicated each partner's performance times measured in stage 1. Before each work session, couples had a free discussion time to plan their work and decide on their strategy. Paper and pens were provided on their table if needed during this bargaining time. At this stage, the experimenters were vigilant to stand back in order to avoid any bias coming from the observer. They remained apart from the process, except to give the starting signal and activate the stopwatch at the beginning of each round.

In these joint work sessions, couples were faced with the problem (1) described in the theoretical section 2.1. The partners had to complete together the same two tasks as in stage 1 in a limited time T of 7 minutes. The required vector of output $(\bar{q}_1, \bar{q}_2)$ was 11 pairs of socks and 5 'T' in the gendered condition, and 5 envelopes and 5 nominative forms in the neutral condition. The time allotted was calibrated so that the two tasks could be fully accomplished by one partner alone, in accordance with the assumptions made in the theoretical model.[25] Each partner could stop working at any time by announcing "free time" to the experimenter, knowing that the free time saved on his/her endowment would be paid to the couple at the individualized wage rate communicated at the beginning of stage 2. The female and male wage rates to be applied were reminded to the couple by the experimenter at the beginning of each work session. Partners were informed at the time of instructions that the free time was indivisible in the sense that when a partner announced "free time", he or she could not return to work. That is, the free-time could be taken only once and for all by each partner in each work session.[26]

The rules of joint production were clearly explained to the participants. Couples were informed that they could choose to share tasks as they wished. They could perform one task each, each of them being paid at his/her wage rate for the free time saved on the time allotted. They could instead choose to let one of them complete the two tasks alone, the

[25]None of the couples in the sample failed to complete the tasks in the time allotted.

[26]There was no reference to market work, domestic work or leisure in the instructions, only the term "free time" was used.

partner performing both tasks being paid at his/her wage rate for the remaining time, and the other at his/her wage rate for his/her whole time endowment. They could also choose any intermediary solution: both accomplishing a share of both tasks, or one working on both tasks and the other on one task only. The experimenter took note of the working time of each partner, as well as their share in the total output for each task.

The design of stage 2 was intended to induce couple members, when working together, to allocate their time according to their comparative advantage and hence maximize their joint payoff under different wage schemes. To illustrate, let us consider two partners whose individual performances on tasks, measured in stage 1, are those described in the upper part of Table 4. Compared to her partner, the woman spent more time on the sock task but less on the bracket task. The individualized wage rates provided to the couple, at the beginning of stage 2, for the three successive joint work sessions are indicated in the middle part of Table 4.

Table 4: An Illustrative Example (Gendered condition)

Performance of subjects in individual session on:	Female time t_i^f	Male time t_i^f	Female relative productivity α_i^f/α_i^m
Task $i = 1$ (Socks)	210 sec.	150 sec.	0.7
Task $i = 2$ (Brackets)	135 sec.	210 sec.	1.5
Wage regimes in couple sessions (in ECU/mn)	Female wage w^f	Male wage w^m	Female relative wage w^f/w^m
1- MWS	300	300	1.0
2- MDE	404	196	2.1
3- WDE	106	494	0.2
Total payoff if the task allocation is:	(in ECU)	(in €)	
Fully efficient in each regime	2 418	69	
Fully stereotyped in each regime	2100	60	

In this example, the woman has a comparative advantage in the bracket task as compared to her partner: her relative productivity in the sock task, α_1^f/α_1^m, given by the inverse ratio of times t_1^m/t_1^f, amounts to .7 and is lower than her relative productivity in the bracket task,

α_2^f/α_2^m, which is about 1.5. Under the first wage scheme, the wage ratio w^f/w^m is set at 1 and hence the opportunity cost of time is the same for both spouses. According to our model (see conditions in Appendix Table A.1), efficiency then dictates that each spouse performs alone the task corresponding to his/her comparative advantage: the woman works on the bracket task and the man on the sock task (MWS regime). This would allow both partners to save a maximum of free time, paid at the same rate of ECU300 per minute, and hence to maximize their joint payoff. Under the second wage scheme, the opportunity cost of time is higher for the woman since her wage rate is more than twice that of her partner, and her wage advantage is also higher than her productivity advantage in the bracket task. Efficiency then requires that the man alone takes in charge both tasks while the woman saves all her time endowment (WDE regime). Finally, under the third wage scheme, the opportunity cost of the woman's time is lower than that of the man since her relative wage rate is about .2, and her relative advantage in the bracket task in not sufficient to compensate for her lower relative wage rate. Efficiency then requires that the woman carries out alone both tasks so that her partner saves all his time endowment (MDE regime). If the couple behaved efficiently in the three regimes, and if the individual performances of the partners were invariant between stages 1 and 3, the couple's total payoff at the end of the experiment would amount to €69, as shown in the lower part of Table 4.[27] If, conversely, the partners systematically shared tasks according to gender roles and stereotypes, that is the man tightened the brackets and the woman sorted the socks in the three joint work sessions, the couple's total earnings would amount to €60. Hence, compared to a fully stereotyped allocation, efficiency raises the couple's gains by more than 15 percent in this example.

[27]The couple's total earning was determined as the average payoff over the 5 work sessions of the experiment: the two individual work sessions of stages 1 and 3 and the three couple work sessions of stage 2. More productive couples could earn more than €100.

4 Results

We begin by displaying summary statistics of behaviors observed during individuals' and couples' work sessions to display the main trends emerging in terms of productivity, differences between treatments (especially gendered/neutral) and between wage schemes. Then we present the results related to theoretical predictions 1 to 4. Descriptive statistics of the variables used in the regression analysis carried out in this section are presented in Appendix D.

4.1 Raw data

Table 5 shows the average individuals' performance times on tasks in gendered and neutral conditions measured at the beginning and at the end of the experiment (stages 1 and 3), as well as ratios of female-to-male productivities (FRP).

Table 5: Summary Statistics of Behavior during Individual Working Sessions

	Gendered condition				Neutral condition			
	Sock Task				**Envelop Task**			
	Measure 1	Measure 2	Variation	N	Measure 1	Measure 2	Variation	N
Female time (sec.)	133 (37.5)	112 (43.0)	-11.10% (33.40)	40	113 (28.9)	104 (28.2)	-7.23% (15.27)	41
Male time (sec.)	157 (37.7)	138 (41.8)	-11.56% (18.66)	40	127 (24.5)	119 (22.7)	-4.72% (18.76)	41
FRP	1.25 (0.39)	1.21 (0.43)	+ 2.29% (0.37)	40	1.18 (0.36)	1.21 (0.38)	+4.76% (0.24)	41
	Bracket Task				**Phone Task**			
	Measure 1	Measure 2	Variation	N	Measure 1	Measure 2	Variation	
Female time (sec.)	145 (26.3)	118 (36.3)	-16.35% (22.80)	40	100 (33.9)	96 (31.8)	-2.50% (19.44)	41
Male time (sec.)	156 (42.0)	133 (42.4)	-14.02% (14.24)	40	105 (23.4)	102 (27.4)	-2.37% (18.58)	41
FRP	1.10 (0.29)	1.09 (0.36)	- 1.12% (0.22)	40	1.11 (0.30)	1.11 (0.27)	+3.04% (0.24)	41

Notes: Standard deviations in parenthesis. FRP>1 means that the woman is more productive than her partner in the task.

Two remarks are worth mentioning here. Firstly, on average, women did better than men in all tasks, both gendered and neutral (they always spent less time performing them), which was rather unexpected, especially concerning the "masculine task" in the gendered condition. However, they performed particularly well in the sock task compared to men. Secondly, there is evidence of learning-by-doing between the first and the second measures of productivity, with a higher increase in productivity for the gendered tasks. These productivity gains,

however, are very similar for men and women, and hence their relative productivity in tasks show very little variation.

Table 6 displays the average characteristics of couples' work sessions. The first part shows the average wages attributed to the female and male partners in each wage scheme, in neutral and gendered sessions, as well as the female-to-male wage ratios. The second part displays the female average contribution to the joint production in each specialization pattern. Finally, the last part demonstrates the couple's time allocation to paid-time and the corresponding payoffs.

Table 6: Summary Statistics of Behavior during Couples' Working Sessions

Wage regime	MDE		MWS		WDE	
Treatment condition	G	N	G	N	G	N
Average Wages*						
Female wage (ECU/mn)	387	377	315	313	188	195
	(25)	(41)	(32)	(35)	(64)	(76)
Male wage (ECU/mn)	213	223	285	287	412	405
	(25)	(41)	(32)	(35)	(64)	(76)
Female wage ratio	1.86	1.76	1.13	1.12	0.49	0.55
	(0.35)	(0.39)	(0.26)	(0.27)	(0.25)	(0.39)
Average female share in:						
– production of task 1 (percent)	56.8	37.5	70.4	59.0	77.4	84.9
	(0.46)	(0.45)	(0.43)	(0.49)	(0.38)	(0.34)
– production of task 2 (percent)	25.0	37.0	33.0	41.5	55.4	69.3
	(0.39)	(0.47)	(0.47)	(0.50)	(0.47)	(0.44)
– global production (percent)	40.4	37.2	51.7	50.2	66.4	77.1
	(0.22)	(0.23)	(0.12)	(0.02)	(0.24)	(0.26)
Payoffs						
Average female paid time (in sec.)	311	346	277	318	238	256
	(58)	(50)	(48)	(28)	(68)	(72)
Average male paid time (in sec.)	250.5	274	281	315	322	367
	(69)	(68)	(38)	(32)	(72)	(61)
Couple's average payoff (ECU)	2903	3199	2797	3179	2987	3418
	(326)	(301)	(313)	(236)	(442)	(283)
Observations	40	41	40	41	40	41

Notes: Standard deviations in parentheses. * Experimentally controlled.

We must remember that in the MDE and WDE regimes, productive efficiency – that is the maximization of the couple's money payoff – dictates that the woman's share in joint

production be 0 and 100 percent respectively. On average, as shown in Table 6, couples' actual choices were far from reaching these efficient values in both regimes, especially in the MDE case (a point to which we shall return). These deviations from efficiency will be econometrically confirmed in the following section.

4.2 Efficiency

The insight behind the experimental design was that, because of the influence of some social effects related to gender, the gender condition would exacerbate productive inefficiencies within the couple. This is exactly what we observed.

4.2.1 Test of Prediction 1

Table 7 displays the two measures of inefficiency, $\bar{d}_1$ and $\bar{d}_2$ (defined in section 2.2), under the neutral and gendered conditions, for the three couples' work sessions of stage 2, where couples faced successively, but in random order, the different wage schemes MWS, WDE and MDE. Remember that $\bar{d}_1$ represents the average proportion of efficient couples, and $\bar{d}_2$ the average distance of actual couples' choice from full efficiency.

On average, couples behaved efficiently in 47.5 percent of the cases for neutral tasks, compared with 33.6 percent for gendered tasks. This difference in couples' behavior between neutral and gendered tasks appears systematic, though it varies to some extent, depending on the wage scheme. In the MWS case, efficiency dictates that each partner specializes entirely in the task corresponding to his/her comparative advantage. This is what most couples have actually done, whatever the nature of the task, particularly under the neutral condition with 73 percent of efficiency, compared with only 60 percent under the gendered condition. By contrast, for the MDE and WDE patterns, couples were more often inefficient than efficient, whatever the nature of the tasks, and even more clearly in the MDE regime where efficiency dictated that the man performed both tasks alone. This could reflect an aversion of partners

Table 7: Impact of the Gendered Treatment on Couple Efficiency

	Proportion of efficient couples $\bar{d}_1$		Distance to efficiency $\bar{d}_2$	
Treatment condition	Neutral (N)	Gendered (G)	Neutral (N)	Gendered (G)
MWS (N=81)	73.2%	60.0%	0.493 (0.867)	0.532 (0.809)
MDE (N=80)	20.0%	15.0%	0.745 (0.451)	0.818 (0.449)
WDE (N=80)	48.8%	25.6%	0.458 (0.512)	0.672 (0.473)
All (N=241)	47.5%	33.6%	0.564 (0.646)	0.674 (0.607)
Mann-Whitney test (p-value)	$H_0 : d_1(N) = d_1(G)$ 0.0281^*		$H_0 : d_2(N) = d_2(G)$ $0.0810^\dagger$	
Test of homogeneity (Mantel-Haenszel)	$\chi^2(2) = 0.83$		$\Pr > \chi^2 = 0.6616$	

Notes: † a significant difference at the 10% level, * at the 5% level and ** at the 1% level.

to extreme inequality within the couple,[28] or a zero aversion – an aversion to do nothing while the partner does everything – which is particularly high when the man is supposed to do everything. This last observation could result from the fact that all the tasks were perceived as "domestic" by participants, hence partners could think inappropriate that men perform them all. However, on average and across the three wage schemes, the proportion of efficient choices was lower, and the distance to efficiency higher, under the gendered condition compared to the neutral condition.

We further tested whether couples assigned to the gendered treatment demonstrated a significantly different behavior in terms of efficiency compared to those assigned to the neutral treatment. Results are reported at the bottom of Table 7. We firstly used the Mantel-Haenszel (1959) test[29] to check whether the effect of the gendered treatment was heterogeneous or could be considered as homogenous across the MWS, MDE and WDE

[28]Note, however, that as participants were informed of the three wage settings from the beginning of stage 2, they also knew that they would have the opportunity to reverse the roles in another session of joint work: this design was used precisely to avoid equity problems.

[29]This test checks the equality of the odds-ratio of the effect of the stereotyped treatment on efficiency in a logistic model, across these different groups (MWS, MDE and WDE).

regimes. The test failed to reject differences between regimes at the 5 percent level, with a p-value of 0.6616, so we chose to implement the tests of the gendered treatment effect by pooling the 3 rounds of joint work. On the pooled data, we implemented a Mann-Whitney-Wilcoxon test[30] (Mann and Whitney 1947; Wilcoxon 1945) which confirms that:

i) the effect of the gendered treatment on the proportion of efficient couples $\bar{d}_1$ is significant and negative (the p-value is 0.0281 and leads to the rejection of equality of means at the 5 percent level).

ii) the distance from efficiency $\bar{d}_2$ is significantly higher under the gendered condition (the p-value equals to 0.0810 and leads to the rejection of the equality of means at the 10 percent level). Hence, inefficiencies in the sharing of tasks between partners are exacerbated when couples have to perform gendered tasks compared to neutral tasks.

Notice that many reasons may prevent members of a couple from reaching full efficiency when they share tasks, such as informational asymmetries regarding their productivity in tasks, a lack of rationality, an individual preference for the tasks they perform poorly, or even a strategic behavior linked to the way the payment is computed and paid. However, there is no reason to believe that these factors would affect neutral and gendered sessions differently. We nevertheless test the robustness of these results in a multivariate setting, taking into account the fact that the error terms for the same couple in the different wage regimes could be correlated. We ran a logistic clustered[31] regression to estimate the effect of the gendered treatment (G equals 1 for the gendered condition and 0 for the neutral condition) on the efficiency of couples' choices, measured by the indicator variable d_1, controlling for wage schemes, geographical location of the experimental session and the experimenter's gender. The results reported in the first panel of Table 8 confirm our previous finding: the probability for a couple to be efficient is significantly reduced by the gendered treatment and is higher

[30]Mann-Whitney U test is the non-parametric alternative test to the independent sample t-test used to test whether two sample means that come from the same population are equal or not.

[31]To control for the fact that the three joint work sessions of a same couple are not independent.

in the MWS and WDE wage regimes compared to the MDE regime. The location of the experimental session and the experimenter's gender have no significant effect on couples behavior. To test the effect of the gendered treatment on the distance from efficiency d_2, linear regression is not adapted because d_2 presents mass points at values 0 and 1 which represent 70 percent of the data. So we chose to run a clustered ordered logit regression using five ordered classes from the lowest distance to the highest distance. Estimation results are presented in the second panel of Table 8. They are very similar, with a significant positive effect of the gendered treatment on the distance to efficiency, a similar effect of the different regimes and no significant effect of the location and experimenter gender dummies. Hence:

Result 1. *Having to perform gender stereotyped tasks reduces the likeliness of choosing an efficient division of labor and increases the distance to efficiency.*

Table 8: Effects of the Gendered Treatment on Couples' Efficiency

	% of Efficicency d_1		Distance to efficiency d_2	
Variable	**Coef.**	**(SD)**	**Coef.**	**(SD)**
G	-0.816*	(0.352)	0.451†	(0.265)
MWS	2.309**	(0.392)	-1.448**	(0.409)
WDE	1.058**	(0.308)	-0.699**	(0.187)
Cergy	0.344	(0.339)	0.002	(0.272)
Male Experimenter	-0.036	(0.335)	-0.062	(0.251)
Intercept	-1.357**	(0.362)		
N	241		241	
Log-likelihood	-138.092		-319.967	

Notes: significance levels are † 10%, ∗ 5%, ∗∗ 1%. Reference category: neutral treatment condition, MDE regime, Paris, female expermimenter.

4.2.2 Test of Prediction 2

We now examine if the gendered treatment exacerbates inefficiencies in the direction which is expected from gender roles and stereotypes. In Table 9, in the gendered condition, the distance to the efficient allocation has been characterized using the expected/unexpected indicators defined in section 2.2: a deviation is considered as "expected" if, compared to

the efficient choice, the woman is actually too much involved in the feminine task (the sock task) and/or the man in the masculine task (the bracket task), while it is considered as "unexpected" otherwise. We compare the results to a placebo case which is obtained in the same way, in the neutral condition, by considering as expected a deviation going in the direction of the task that is on average preferred by women.[32]

Table 9: Effect of Gender Roles and/or Stereotypes on Efficiency

Treatment	Distance d_2	Expected d_2^{exp}	Unexpected d_2^{unexp}	Mann-Whitney test $d_2^{exp} = d_2^{unexp}$
Gendered	0.67 (0.61)	0.46 (0.58)	0.22 (0.44)	p-value= 0.0022**
Neutral (placebo)	0.56 (0.65)	0.31 (0.53)	0.26 (0.52)	p-value=0.3507

Notes: SD in parentheses. ** significant at the 1% level.

The results of this comparison, reported in Table 9, demonstrate that in the neutral condition, the difference between the expected and unexpected deviations, respectively d_2^{exp} and d_2^{unexp} is not significant. By contrast, in the gendered condition, the test shows a positive significant difference between the two directions: couples did deviate more in the stereotyped direction, confirming our prediction 2. Clearly, for couples, the stereotyped behavior has power to attract them away from efficiency.

Result 2. *When a deviation from efficiency occurs, it goes usually in the direction of the stereotype.*

4.3 Investigating the Sources of Inefficiency

Now, we explore the respective roles of direct individual preferences for tasks, gendered social norms and gendered stereotypes (in the sense of false beliefs concerning individual productivities) in explaining the "excess" inefficiency of couples' choices observed under the gendered condition compared to the neutral condition.

[32]Individual preferences for tasks were elicited from questionnaires (Table 10). They show that regarding neutral tasks, men preferred the phone task (task 2) and women the envelope task (task 1).

4.3.1 Differences in Task Preferences

Preferences differ between tasks for the same individual but also differ between men and women for the same task. Moreover, in our experimental design, tasks themselves differ between treatments. The econometric tests below aim at taking into account these different dimensions.

Preferences for tasks were elicited using individual anonymous questionnaires filled in by the subjects during the course of the experiment. In questionnaire Q1, filled in at the end of the first individual work session and before the joint work sessions, the subjects were asked the following question: "If you could choose to perform only one of the two tasks proposed, which one would you choose". Furthermore, in order to get information about the intensity of preferences, at the end of the experiment, in questionnaire Q3, the participants were asked: "did you find each of these tasks – very pleasant, – rather pleasant, – rather unpleasant, – very unpleasant".

Table 10 presents the answers to questionnaire Q1. On average, in gendered sessions women preferred the sock task, while men preferred the bracket task, confirming the expected gender stereotype and/or habits. In neutral sessions, women preferred the envelope task, while men preferred the phone task. Surprisingly, gender differences in preferences happened to be particularly large with regard to neutral tasks. Indeed, the difference between the proportion of men and women who favor the phone task reaches 38.4 percentage points, whereas the difference between the proportion of woman and men who favor the sock task is 30, a result which was not salient in pilot sessions. This unexpected result however gives us a first insight into the role of task preferences on couples' choices. The difference in preferences being on average larger for neutral tasks than for gendered tasks, if preferences had a major impact on deviations from efficiency, the average observed deviation should also have been wider for neutral than for gendered tasks. Yet, as we saw, exactly the opposite happened.

To go further, we used the answers to questionnaire Q3 designed to measure the intensity of preferences. We ran several regressions using d_1 (clustered logits) and d_2 (clustered ordered

Table 10: Preferred Tasks by Gender

Neutral condition	Phone No	Envelopes
Women (N=41)	48.8 %	51.2 %
Men (N=39)	87.2 %	12.8 %
Gendered condition	Socks	Brackets
Women (N=40)	72.5 %	27.5 %
Men (N=40)	42.5 %	57.5 %

logits) as the dependent variable and the treatment dummy G, and variables measuring the differences in the intensity of preferences between the two tasks for men and for women (DifT1T2M and DifT1T2W respectively), or the differences in the intensity of preferences for tasks between genders (DifMWT1 and DifMWT2 respectively) as independent variables.[33] Estimation results are reported in Table 11. The fact that men or women prefer one task over the other, or that they do not have the same preferences for tasks, has no significant influence on the probability for a couple to make an efficient choice (d_1) or on the distance to efficiency (d_2). However, once preferences are controlled for, the gendered treatment keeps its strong negative effect on efficiency: the probability for a couple to be efficient is reduced and the distance to efficiency increases. These results confirm our Prediction 3 – individuals' preferences for tasks did not play a significant role in the allocation of tasks within couples – and hence the standard assumption made in economic models of the household regarding the separability of household production decisions from preferences.

Another way of examining the role of preferences is to compare between treatments the effect of a change in relative wages on the couples' allocation of tasks. Given that preferences for tasks of couples' members remain constant over work sessions, if couples re-allocated tasks more efficiently following a change in the wage scheme under the neutral condition than under the gendered condition, then this will be an indication that the role of preferences in observed inefficiencies is not prominent. Table 12 displays a fixed effect panel estimation of d_2. The

[33]The answers to questionnaire Q3 being ranked for each task from 1 (for very unpleasant) to 4 (for very pleasant), the values of the preference difference variables range between -3 to +3.

Table 11: The Effect of Task Preferences on Efficiency

	% of Efficiency d_1				Distance to efficiency d_2			
	(1)		**(2)**		**(1)**		**(2)**	
Variable	**Coef.**	**(SD)**	**Coef.**	**(SD)**	**Coef.**	**(SD)**	**Coef.**	**(SD)**
G	-0.652*	(0.275)	-0.642*	(0.280)	0.398	(0.247)	0.449†	(0.246)
DifT1T2M	-0.054	(0.174)			-0.079	(0.187)		
DifT1T2W	-0.185	(0.193)			0.083	(0.124)		
DifMWT1			0.126	(0.186)			-0.054	(0.162)
DifMWT2			0.180	(0.170)			-0.179	(0.145)
Constant	-0.011	(0.206)	-0.085	(0.181)				
N	241		241		241		241	
Log-likelihood	-159.757		-159.092		-330.639		-329.723	

Notes: significance levels are † 10% ∗ 5% ∗∗: 1%. Ref. category: neutral condition (G=0).

independent variable is the change of wage regime, from the MDE regime where the male relative wage rate is the lowest, to the intermediary regime MWS and then to the WDE regime where the male-to-female wage ratio is the highest. What appears clearly in Table 12 is that couples reallocate tasks in a significantly efficient way in response to wage changes under the neutral condition, while this is not the case under the gendered condition. Hence,

Result 3. *Gender differences in preferences for tasks cannot explain the efficiency gap between the gendered and the neutral conditions.*

Table 12: Fixed Effect Panel Estimation of d_2

	Gendered		**Neutral**	
Variable	**Coef.**	**(SD)**	**Coef.**	**(SD)**
Wage_regime_Panel	-0.074	(0.066)	-0.143*	(0.069)
Intercept	0.821**	(0.143)	0.851**	(0.150)
N	119		122	
R^2	0.016		0.051	

Notes: significance levels are † 10% ∗ 5% ∗∗ 1%.

4.3.2 Gender Stereotypes : the Role of Beliefs and Information

As preferences do not appear to contribute significantly to the observed inefficiencies, it remains to try to disentangle the effect of gender social norms from that of stereotypes. The informational treatment was designed for this purpose. If social norms were the main explanation of the observed efficiency gap between the gendered and neutral treatments, then provide couples with information about the productivity in tasks of their members should not really make any difference. Indeed, social norms weight on individual or group behavior because it is utility costly[34] for individuals to deviate from social prescriptions. Couples may rationally decide to deviate from efficiency and accept the resulting money loss to avoid the discomfort of the violation of the gender codes of behavior socially prescribed. The issue in this case is not related to the lack of information. By contrast, if deviations from efficiency were mainly due to false beliefs about the relative productivity of men and women in tasks, induced by gender stereotypes, then revealing this information should lead couples to revise their choices in the direction of efficiency, and hence reduce the probability of observing stereotyped behaviors (Prediction 4).

We have seen previously that, on average, women had an absolute advantage in all tasks. A review of partners' comparative advantages, computed from individual performances measured in the first stage of the experiment, indicates that women, compared to men, were more likely to have a comparative advantage in the envelope task (in 58.5 percent of the observations) than in the phone task, and in the sock task than in the bracket task (60 percent of the observations).

Let us now examine the spouses' beliefs about these performances, which were collected via Questionnaire 1. Table 13 reports the actual relative performances of women in tasks compared to their partner, as well as the beliefs that she and her partner had about them. Concerning gendered tasks, partners' beliefs confirm our assumption concerning the gendered nature of the sock and bracket tasks and are in accordance with gender stereotypes: 85 per-

[34] Violating the prescriptions evokes anxiety and discomfort in oneself and in others.

Table 13: Female Relative Performances: Beliefs and Observations

Treatment	Task	Beliefs according to	Woman at least as fast as partner (%)	Man at least as fast as partner (%)
Neutral	Phone	Women	65.8	56.1
(N=41)		Men	56.4	69.2
		Observed	61.0	48.8
	Envelopes	Women	73.2	51.2
		Men	76.9	56.4
		Observed	70.7	43.9
Gendered	Socks	Women	84.6	43.6
(N=40)		Men	85.0	52.5
		Observed	77.5	35.0
	Brackets	Women	35.0	82.5
		Men	47.5	75.0
		Observed	65.0	52.5

cent of women and of men believed that in their couple, the woman would be a least as fast[35] as her partner in the sock task, with about 50 percent of them thinking that the man would be. Note however that for socks, beliefs were not so far from observed performances. 77.5 percent of women were actually at leat as fast as their partner, compared to 35 percent for men. Beliefs were also in line with the stereotype for the bracket task, but wrongly in this case: only 35 percent of women and 47.5 percent of men believed that the woman would be at least as fast as her partner on brackets, though the observed figure is 65 percent. Hence, beliefs about gendered tasks seemed clearly driven by gender stereotypes. Regarding neutral tasks, beliefs appear gender differentiated too, but to a lesser extent. Concerning the phone task, each partner tended to believe that he/she would do at least as well as his/her partner (56.4 percent of men, 65.8 percent of women). In fact, women performed better (61 percent were at least as fast). Note, however that the phone task is that in which men had more often a comparative advantage. The envelope task seems to have been

[35]The threshold for equality was set at +/-5 percent.

perceived by both partners (73 percent of women, 77 percent of men) as a more "feminine" task, but rightly this time, since 71 percent of women did perform at least as well as their partner on that task. Note that, generally, men tended to have a better opinion of their own performances in tasks than that expressed by their partner (except for the bracket task where women strongly overestimate men's productivity). Moreover, rightly, women tended to have a better opinion of their own performances than of those of their partner (again except for brackets). Finally, men more systematically and strongly overestimated their own performance than women. This confirms the experimental literature on the "overconfidence" of men (see, among others, Niederle and Vesterlund 2007). Interestingly, women tended also to overestimate their partner's performance (except for the phone task) and hence seem to adhere more strongly than their companion to stereotypes.

We now examine the effect on couples' decisions of the informational treatment, which consisted in providing or not couples with information on their individual productivities in tasks, at the time of the joint work sessions. Table 14 reports the difference between the average deviation to efficiency observed under the informed and non-informed conditions. The informational treatment has globally a modest and non significant effect on the proportion of efficient outcomes ($\bar{d}_1$) in both conditions. But interestingly enough, information tends to reduce the distance to efficiency ($\bar{d}_2$), and more particularly the expected distance ($\bar{d}_2^{exp}$), that is deviations toward the stereotyped allocation of tasks.

Table 14: Effect of the Informational Treatment

	% of efficient couples $\bar{d}_1$	Distance to efficiency $\bar{d}_2$	Expected distance $\bar{d}_2^{exp}$	Unexp. distance $\bar{d}_2^{unexp}$
Treatment conditions				
Informed	40.52%	0.596	0.358	0.244
(N=116)	(0.49)	(0.60)	(0.52)	(0.48)
Not Informed	40.80%	0.640	0.548	0.190
(N=125)	(0.49)	(0.65)	(0.63)	(0.40)
Mann-Whitney test[1]				
P-value	0.9645	0.7576	0.1177	0.6260

Notes: significance levels are † 10% ∗ 5% ∗∗ 1%.

To assess the effect of information and beliefs, we now focus on the impact of mistaken beliefs. We ran two regressions on d_1 and d_2 (a clustered logit for d_1 and a clustered ordered logit for d_2) with, as independent variable, the dummy variable 'False beliefs' (thats is beliefs that are inverse to the true comparative advantages) interacted with the informational treatment dummy (that equals 1 under the informed treatment condition). The gendered treatment dummy and wage regime dummies are added as control variables. The results are displayed respectively in panels 1 and 2 of Table 15. They show that mistaken beliefs signifi-

Table 15: Effect of Information Provision on Couples' Efficiency

	% of efficiency d_1		Distance to efficiency d_2	
Variable	**Coef.**	**(SD)**	**Coef.**	**(SD)**
Not informed×False beliefs	-1.745**	(0.541)	0.808*	(0.320)
Informed×True beliefs	-0.350	(0.383)	0.132	(0.318)
Informed×Faulse belief	-1.180**	(0.332)	0.575*	(0.261)
G	-0.558	(0.346)	0.395	(0.250)
MDE	2.449**	(0.429)	-1.420**	(0.415)
WDE	1.101**	(0.323)	-0.733**	(0.198)
Intercept	-0.897**	(0.315)		
N	241		241	
Log-likelihood	-131.173		-316.888	

Notes: significance levels are † 10% * 5% ** 1%. Ref. category: neutral condition, not informed, true beliefs.

cantly increase inefficiency. First, notice that the effect of the gendered treatment is still negative on d_1 and positive on d_2 but both coefficients become non-significant. Looking at interaction terms, we see that inefficiencies increase strongly and significantly when beliefs are mistaken. As could be expected, information plays no role when beliefs are consistent with comparative advantages. Conversely, information does lower inefficiency (the coefficient of the dummy 'False beliefs' decreases in both estimations when information is provided) but the difference between the coefficients remains non significant. Note that all couples who had false beliefs in our sample, wrongly believed that the woman had the comparative advantage in the sock task while she had it in the bracket task (their beliefs were indeed

always attracted by the stereotype). Moreover, providing information to them resulted in a more efficient sharing of tasks in the MWS regime, where this information is the more likely to change behavior (see Table D.2 in Appendix D, noting that the small number of observations in this case leads to remain cautious on this result). Thus, the econometric analysis conducted confirm Prediction 4:

Result 4. *Provide couples information about the individual performances on tasks of each partner partly reduces the stereotype effect.*

Globally, if information provision about partners' productivities on tasks has no significant effect on the proportion of efficient outcomes, it significantly impacts couples' behavior under the gender condition but not under the neutral condition. This result shows that the treatment of information differs within couples depending on the nature of the tasks they have to perform. Private information regarding productivity on tasks seem to flow more fluently between partners within couple as far as neutral tasks are concerned, while it seems impeded by stereotypes. There is no doubt, given what we have observed in the course of the experiment, that the gendered treatment induced a negotiation shortcut in the couple decision-making process.[36] So, decision-making within couples could well take paths that are difficult to describe in current economic household models. Our observations are in line with researches in social psychology which recognize the role of social norms and stereotypes as an energy and time-saving device (see e.g. Macrae et al. 1994).

5 Conclusion

The design of the present experiment aimed at testing the efficiency of task sharing within couple, especially when partners are performing "gendered tasks". Eighty one established couples participated in a lab experiment where they had to work individually and jointly on two concrete day-to-day tasks. After having measured the productivity on tasks of each

[36]We clearly noticed during experimental sessions that partners exchanged more and computed more in neutral sessions than in gendered sessions where the bargaining time was much shorter.

partner separately, we let them choose together how to carry out jointly both tasks, in a limited time. Traditional economic theory predicts that couples should allocate their time according to their comparative advantage. Our experimental results show that having to perform gendered tasks induce couples to deviate from efficiency most of the time. Couples significantly deviate more often, and in a larger extent, when they face gendered tasks as compared to neutral tasks. They also deviate more often in the direction expected from gender stereotypes: the woman taking in charge the "feminine task" and the man the "masculine" task.

Our results also provide further explanations of the results. Three main causes for such deviations are provided by theory: either preferences directly entering the production function, or social norms (there is a utility cost to deviate from "gender roles"), or finally stereotypes (false beliefs about the female-male relative productivity in "gendered tasks"). The results of our experiment make it possible to exclude the first factor, i.e. preferences, as the main cause of couples' productive inefficiency. Moreover, even if our results do not allow to fully disentangle the effects of gender roles from those of gender stereotypes in explaining inefficiencies, we show that information matters. Our results indeed indicate that false beliefs based on stereotypes about relative productivities of men and women in gendered tasks significantly increase the distance between the allocation chosen by couples and the efficient allocation in the direction expected from the stereotype: women on average overspecialize in the "feminine task" and men the in masculine "task". Moreover, providing information tends to decrease the impact of false beliefs. Social representations related to gender influence the household decision-making process, alongside the usual factors considered by economists, and should certainly be better taken into account in order to better understand the economic decisions of households.

Appendix A. Model resolution

The couple will behave efficiently if the partners choose a division of labor which maximizes their joint gain, that is, solves the following program (1):

$$\begin{cases} \max w^f(T - t_1^f - t_2^f) + w^m(T - t_1^m - t_2^m) \\ \text{s.t. } \alpha_1^f t_1^f + \alpha_1^m t_1^m \geq \bar{q}_1 & (\lambda_1 \geq 0) \\ \quad \alpha_2^f t_2^f + \alpha_2^m t_2^m \geq \bar{q}_2 & (\lambda_2 \geq 0) \\ \quad t_1^f, t_2^f, t_1^m, t_2^m \geq 0 \end{cases}$$

We restrict our attention to the cases where the man or the woman alone has enough time to perform entirely the two domestic tasks, that is, we assume: $\frac{\bar{q}_1}{\alpha_1^f} + \frac{\bar{q}_2}{\alpha_2^f} < T$ and $\frac{\bar{q}_1}{\alpha_1^m} + \frac{\bar{q}_2}{\alpha_2^m} < T$. The Kuhn-Tucker conditions of program 1 are:

$$\begin{cases} t_1^f \geq 0 \quad -w^f + \alpha_1^f \leq \lambda_1 \quad \text{and } t_1^f(-w^f + \alpha_1^f \lambda_1) = 0 \\ t_2^f \geq 0 \quad -w^f + \alpha_2^f \leq \lambda_2 \quad \text{and } t_f^2(-w^f + \alpha_1^f \lambda_2) = 0 \\ t_1^m \geq 0 \quad -w^m + \alpha_1^m \leq \lambda_1 \quad \text{and } t_1^m(-w^m + \alpha_1^m \lambda_1) = 0 \\ t_2^m \geq 0 \quad -w^m + \alpha_2^m \leq \lambda_2 \quad \text{and } t_m^2(-w^m + \alpha_1^m \lambda_2) = 0 \end{cases}$$

These conditions define 9 possible equilibria and 4 specialization patterns, depending on the value of productivity parameters and wage rates, as indicated in Table A.1. Note that the fourth regime correspond to limiting cases which are not relevant experimentally.

Table A.1: Equilibrium Specialization Patterns

Specialization Regime	Conditions on parameters	Optimal allocation of partners' time			
		t_1^f	t_2^f	t_1^m	t_2^m
WDE	$\frac{w^f}{w^m} < \frac{\alpha_1^f}{\alpha_1^m}$ and $\frac{w^f}{w^m} < \frac{\alpha_2^f}{\alpha_2^m}$	$\frac{\bar{q}_1}{\alpha_1^f}$	$\frac{\bar{q}_2}{\alpha_2^f}$	0	0
MDE	$\frac{w^f}{w^m} > \frac{\alpha_1^f}{\alpha_1^m}$ and $\frac{w^f}{w^m} > \frac{\alpha_2^f}{\alpha_2^m}$	0	0	$\frac{\bar{q}_1}{\alpha_1^m}$	$\frac{\bar{q}_2}{\alpha_2^m}$
MWS	$\frac{\alpha_2^f}{\alpha_2^m} < \frac{w^f}{w^m} < \frac{\alpha_1^f}{\alpha_1^m}$	$\frac{\bar{q}_1}{\alpha_1^f}$	0	0	$\frac{\bar{q}_2}{\alpha_2^m}$
	$\frac{\alpha_1^f}{\alpha_1^m} < \frac{w^f}{w^m} < \frac{\alpha_2^f}{\alpha_2^m}$	0	$\frac{\bar{q}_2}{\alpha_2^f}$	$\frac{\bar{q}_1}{\alpha_1^m}$	0
Fourth regime	$\frac{w^f}{w^m} = \frac{\alpha_1^f}{\alpha_1^m} < \frac{\alpha_2^f}{\alpha_2^m}$	$\frac{\bar{q}_1}{\alpha_1^f} - \frac{\alpha_1^m}{\alpha_1^f} t_1^m$	$\frac{\bar{q}_2}{\alpha_2^f}$	$0 \leq t_1^m \leq \frac{\bar{q}_1}{\alpha_1^m}$	0
	$\frac{w^f}{w^m} = \frac{\alpha_2^f}{\alpha_2^m} < \frac{\alpha_1^f}{\alpha_1^m}$	$\frac{\bar{q}_1}{\alpha_1^f}$	$\frac{\bar{q}_2}{\alpha_2^f} - \frac{\alpha_2^m}{\alpha_2^f} t_2^m$	0	$0 \leq t_2^m \leq \frac{\bar{q}_2}{\alpha_2^m}$
	$\frac{\alpha_2^f}{\alpha_2^m} < \frac{w^f}{w^m} = \frac{\alpha_1^f}{\alpha_1^m}$	$0 \leq t_1^f \leq \frac{\bar{q}_1}{\alpha_1^f}$	0	$\frac{\bar{q}_1}{\alpha_1^m} - \frac{\alpha_1^f}{\alpha_1^m} t_1^f$	$\frac{\bar{q}_2}{\alpha_2^m}$
	$\frac{\alpha_1^f}{\alpha_1^m} < \frac{w^f}{w^m} = \frac{\alpha_2^f}{\alpha_2^m}$	0	$0 \leq t_2^f \leq \frac{\bar{q}_2}{\alpha_2^f}$	$\frac{\bar{q}_1}{\alpha_1^m}$	$\frac{\bar{q}_2}{\alpha_2^m} - \frac{\alpha_2^f}{\alpha_2^m} t_2^f$
	$\frac{w^f}{w^m} = \frac{\alpha_1^f}{\alpha_1^m} = \frac{\alpha_2^f}{\alpha_2^m}$	$0 \leq t_1^f \leq \frac{\bar{q}_1}{\alpha_1^f}$	$0 \leq t_2^f \leq \frac{\bar{q}_2}{\alpha_2^f}$	$\frac{\bar{q}_1}{\alpha_1^m} - \frac{\alpha_1^f}{\alpha_1^m} t_1^f$	$\frac{\bar{q}_2}{\alpha_2^m} - \frac{\alpha_2^f}{\alpha_2^m} t_2^f$

Appendix B. Questionnaires

B1. Individual and Anonymous Questionnaire Q1

N° Anonymat: _____
Date: _____
Session: _____

1. If you could choose to realize only one of the two tasks proposed, which of them would you choose:

   ☐ the task “Phone Numbers”/ “Socks”

   ☐ the task “Envelopes”/ “Brackets”

2. Do you think that you will be:

   ☐ faster on the task “Phone Numbers”/ “Socks”

   ☐ faster on the task “Envelopes”/ “Brackets”

   ☐ equally fast on both tasks

3. In your opinion, compared to you, your partner will be:

on the task “Phone numbers”/ “Socks”:	☐ faster
	☐ less fast
	☐ equally fast
on the task “Envelopes”/ “Brackets”:	☐ faster
	☐ less fast
	☐ equally fast

B2. Individual and Anonymous Questionnaire Q3

N° Anonymat: _____

Date: _____

Session: _____

1. Did you find each realized task:

"Phone numbers"/ "Socks":	☐ Very easy	"Envelopes"/ "Brackets":	☐ Very easy
	☐ Rather easy		☐ Rather easy
	☐ Rather difficult		☐ Rather difficult
	☐ Very difficult		☐ Very difficult

2. Did you find each realized task:

"Phone numbers"/ "Socks":	☐ Very pleasant	"Envelopes"/ "Brackets":	☐ Very pleasant
	☐ Rather pleasant		☐ Ratherpleasant
	☐ Rather unpleasant		☐ Rather unpleasant
	☐ Very unpleasant		☐ Very unpleasant

3. Have you got further remarks concerning the content or the progress of the experience?

_____

_____

_____

_____

Thank you for your participation!

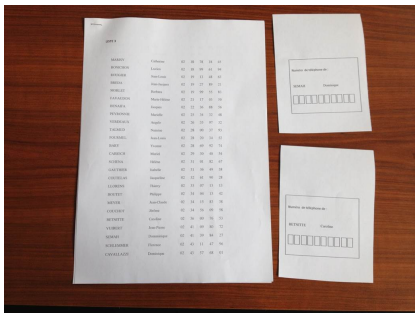
Appendix C. The Tasks

Figure C.1: Gendered Tasks

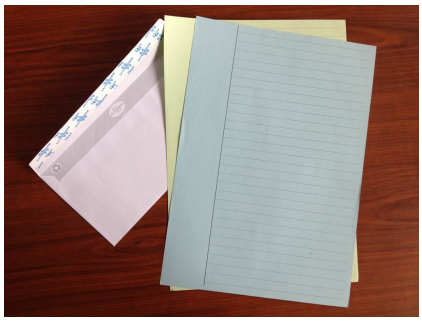
(a) The Sock Task

(b) The Bracket Task

Figure C.2: Neutral Tasks



(a) The Phone Task



(b) The Envelope Task

Appendix D

Table D.1: Summary statistics

Variable	Mean	SD
G	0.494	0.501
Male experimenter	0.481	0.501
DifT1T2M	0.333	0.818
DifT1T2W	0.21	0.750
DifMWT1	-0.037	0.810
DifMWT2	0.086	0.893
Informed	0.481	0.501
False beliefs	0.247	0.432
N	243	

Table D.2: Effect of Information on d_2^{exp} in the MWS Regime:

Variable	Coef.	(SD)
Not informed×False beliefs	3.743**	(1.299)
Informed×True beliefs	0.480	(1.285)
Informed×False beliefs	1.496	(1.562)
N	40	
Log-likelihood	-25.956	

Notes: Ordered logit estimation. Ref. category: not informed, true beliefs.

Significance levels: † 10%, ∗ 5%, ∗∗ 1%.

References

Aguiar, M. and E. Hurst (2007). "Measuring Trends in Leisure: The Allocation of Time over Five Decades". *The Quarterly Journal of Economics* 122 (3), pp. 969–1006.

Akerlof, G. A and R. E. Kranton (2000). "Economics and Identity". *The Quarterly Journal of Economics* 115 (3), pp. 715–753.

Aliaga, C. and K. Winqvist (2003). "How Women and Men Spend their Time. Results from 13 European Countries". *Statistics in Focus Theme 3*. Luxembourg: Eurostat.

Anxo, D., D. Flood, and Y. Kocoglu (2002). "Offre de travail et répartition des activités domestiques et parentales au sein du couple : une comparaison entre la France et la Suède". *Economie et Statistique* 352 (1), pp. 127–150.

Apps, P. F. and R. Rees (1997). "Collective Labor Supply and Household Production". *The Journal of Political Economy* 105 (1), pp. 178–190.

Aronsson, T., S. 0. Daunfeldt, and M. Wikström (2001). "Estimating Intrahousehold Allocation in a Collective Model with Household Production". *Journal of Population Economics* 14, pp. 569–584.

Beblo, M. et al. (2015). "Efficiency-Equality Trade-off within French and German Couples: A Comparative Experimental Study". *Annals of Economics and Statistics* 117-118, pp. 233–252.

Becker, G. S. (1981). *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bertrand, M., E. Kamenica, and J. Pan (2013). "Gender Identity and Relative Income within Households". *The Quarterly Journal of Economics* 130 (2), pp. 571–614.

Bittman, M. et al. (2003). "When Does Gender Trump Money? Bargaining and Time in Household Work". *American Journal of Sociology* 109 (1), pp. 186–214.

Blau, F. and L. Kahn (2006). "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence". *Industrial & Labor Relations Review* 60 (1), pp. 45–66.

Carter, M. and E. Katz (1997). "Separate Spheres and the Conjugal Contract: Understanding the Impact of Gender-Biased Development". *Intrahousehold Resource Allocation in De-*

veloping Countries: Methods, Models and Policies. Ed. by J. Hoddinott L. Haddad and H. Alderman. Baltimore, Md., U.S.A.: Johns Hopkins University Press (for the International Food Policy Research Institute).

Chao, L.W. and H.P Kohler (2007). *The behavioral economics of altruism, reciprocity, and transfers within families and rural communities: Evidence from Sub-Saharan Africa*. Tech. rep. University of Pennsylvania.

Chiappori, P. A. (1997). "Introducing Household Production in Collective Models of Labor Supply". *Journal of Political Economy* 105 (1), pp. 191–209.

Coate, S. and G. C. Loury (1993). "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *The American Economic Review* 83 (5), pp. 1220–40.

Cochard, H. Couprie, and A. Hopfensitz (2016). "Do spouses cooperate ? If not: why ?" *Review of Economics of the Household* 14 (1), pp. 1–26.

Coffman, K. (2014). "Evidence on Self-stereotyping and the Contribution of Ideas". *The Quartely Journal of Economics*, pp. 1625–1660.

Croson, R. and U. Gneezy (2009). "Gender Differences in Preferences". *Journal of Economic Literature* 47 (2), pp. 1–27.

Cudeville, E. and M. Recoules (2015). "Household Behavior and Social Norms: A Conjugal Contract Model". *Annals of Economics and Statistics* (117-118), pp. 279–312.

Duflo, E. and C. Udry (2004). "Intrahousehold Resource Allocation in Côte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices". *NBER Working Papers* (w10498).

Fahr, R. and B. Irlenbusch (2000). "Fairness as a constraint on trust in reciprocity: earned property rights in a reciprocal exchange experiment". *Economics Letters* 66 (3), pp. 275–282.

Falk, A. and A. Ichino (2006). "Clean evidence on peer effects". *Journal of labor economics* 24 (1), pp. 39–57.

Goldschmidt-Clermont, L. and E. Pagnossin-Aligisakis (1995). *Measures of Unrecorded Economic Activities in Fourteen Countries*. New York: UNDP Human Development Report Office.

Hersch, J. and L.S. Stratton (1994). "Housework, Wages, and the Division of Housework Time for Employed Spouses". *The American Economic Review* 84 (2), pp. 120–25.

Kalenkoski, C. M., D. C. Ribar, and L. Stratton (2009). "The Influence of Wages on Parents' Allocations of Time to Child Care and Market Work in the United Kingdom". *Journal of Population Economics* 22 (2), pp. 399–419.

Lundberg, S. and R. A. Pollak (1993). "Separate Spheres Bargaining and the Marriage Market". *Journal of Political Economy* 101 (6), pp. 988–1010.

Macrae, C., A. Milne, and G. Bodenhausen (1994). "Stereotypes as Energy-saving Sevices: a Peek Inside the Cognitive Toolbox". *Journal of Personality and Social Psychology* 66 (1), pp. 37–47.

Mann, H. B. and D. R. Whitney (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other". *Annals of Mathematical Statistics* 18 (1), pp. 50–60.

Munro, A. (2015). *Intra-household Experiments: a survey and some methodological observations*. Tech. rep. National Graduate Institute for Policy Studies.

Munro, A. et al. (2010). *The lion's share. An experimental analysis of polygamy in Northern Nigeria*. Tech. rep. GRIPS Discussion Papers.

— (2014). "Autonomy and efficiency. an experiment on household decisions in two regions of India". *Journal of the Japanese and International Economies* 33, pp. 114–133.

Niederle, M. and L. Vesterlund (2007). "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quaterly Journal of Economics* 122 (3), pp. 1067–1101.

Peters, H. E. et al. (2004). "Free-riding and the provision of public goods in the family: A laboratory experiment". *International Economic Review* 45 (1), pp. 283–299.

Rapoport, B. and C. Sofer (2005). "Pure Production Factors and the Sharing Rule: Estimating Collective Models with Household Production". *TEAM Working Papers*. University of Paris 1.

Régnier-Loilier, A. (2009). "arrivée d'un enfant modifie-t-elle la répartition des tâches domestiques au sein du couple?" *Population & Sociétés* (461).

Rizavi, S. S. and C. Sofer (2010). "Household Division of Labor: Is There Any Escape From Traditional Gender Roles?" *CES Working Paper*. 2010.09. University of Paris 1.

Sevilla-Sanz, A., J. I. Gimenez-Nadal, and C. Fernandez (2010). "Gender Roles and the Division of Unpaid Work in Spanish Households". *Feminist Economics* 16 (4), pp. 137–184.

Sofer, C. and C. Thibout (2011). "Stereotypes upon Abilities in Domestic Production and Household Behaviour". *CES Working Paper*. 2011.75. University of Paris 1.

Udry, C. (1996). "Gender, Agricultural Production, and the Theory of the Household". *Journal of Political Economy* 104 (5), pp. 1010–46.

West, C. and D. H. Zimmerman (1987). "Doing Gender". *Gender and Society* 1 (2), pp. 125–151.

Wilcoxon, F. (1945). "Individual Comparisons by Ranking Methods". *Biometrics Bulletin* (1), pp. 80–83.

Winqvist, K. (2004). *How Europeans spend their timeEveryday life of women and men*. Ed. by Pocketbook. Luxembourg: Office for Official Publications of the European Communities.