UMR 8184
Thema
THéorie Économique, Modélisation et Applications

THEMA Working Paper n°2015-06
Université de Cergy-Pontoise, France

EFFICIENCY VERSUS STEREOTYPES: AN EXPERIMENT IN DOMESTIC PRODUCTION

Hélène Couprie

Elisabeth Cudeville

Catherine Sofer

Février 2015

UNIVERSITÉ de Cergy-Pontoise

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

EFFICIENCY VERSUS STEREOTYPES: AN EXPERIMENT IN DOMESTIC PRODUCTION[*]

Hélène Couprie, [•]
Elisabeth Cudeville[*],
Catherine Sofer[♣]

February 2015

Abstract

Most household models assume that decisions taken inside the family are Pareto optimal. However, empirical studies cast doubts upon the efficiency assumption. The sharing of time among men and women between market work and household work is highly differentiated by gender. In this paper we examine whether couples deviate from efficiency in household production decisions, using an experimental design in which subjects are real couples. The aim of the experiment is to mimic the sharing of highly-gendered household tasks. We compare the sharing of gendered tasks to that of more neutral tasks. By measuring individual productivity in each task, we can see if couples tend to deviate from efficiency, and by how much in each case. As we show that they deviate more when sharing gendered tasks, we also explore why, looking at different possible explanations, and we find evidence of the impact of stereotypes on inefficiencies.

Keywords: Stereotypes, social norms, household production, time allocation, experiment, production function, household behavior, intra-household decision-making.

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

---

[*]We thank P. Apps, P-A Chiappori, and the participants in the Workshop "Economics of Gender " in Nice for useful discussions and comments. Financial support from the French National Research Agency (ANR "Ginhdila") is gratefully acknowledged.

[•] Université de Cergy-Pontoise, email : helene.couprie@gmail.com
[*] Ecole d'Economie de Paris et Université Paris1-Panthéon-Sorbonne, email : elisabeth.cudeville@univ-paris1.fr
[♣] Ecole d'Economie de Paris et Université Paris1-Panthéon-Sorbonne, email:catherine.sofer@univ-paris1.fr; tel : +33 (0)686668769. Contact address: MSE, 106-112 bd de l'Hopital, 75013 Paris, FRANCE

Efficiency versus Stereotypes: an Experiment in Domestic Production

Hélène Couprie, [•]
Elisabeth Cudeville[*],
Catherine Sofer[♣]

Introduction

In many places in the world, gender inequalities in leisure time appear persistently despite the decline in the gender pay gap (Aguiar and Hurst, 2007). The reasons for this are far from being satisfactorily explained by economists. This article presents experimental results that show a major negative impact of stereotyped gender roles on household productive efficiency which could explain part of this persistence.

In most countries, the allocation of time-uses remains highly differentiated by gender (Goldschmidt-Clermont and Pagnossin-Aligisakis, 1995, Miranda, 2011). Women in Europe spend 60% to 70% of their working time on domestic tasks and 30% to 40% in paid-work. European men spend 55% to 65% of their working time in market work, and 35% to 45% in household work. Total working time appears generally higher for women than for men (Winqvist, 2004). In France in 2009, women contributed to approximately 80% of the domestic household tasks (Dumontier *et al.*, 2002). 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

[•] Université de Cergy-Pontoise, email : helene.couprie@gmail.com
[*] Paris School of Economics and Université Paris1-Panthéon-Sorbonne, email : elisabeth .cudeville@univ-paris1.fr
[♣] Paris School of Economics and Université Paris1-Panthéon-Sorbonne, email :catherine.sofer@univ-paris1.fr: contact address : MSE, 106-112 bd de l'Hopital, 75013 Paris, FRANCE

Kahn, 2006) and earned income, through shorter work duration, as well as through flatter careers, could be consequences of the unequal sharing of household work.

Heterogeneity in productivity at home and in the labor market could explain why women's and men's choices could differ. However, it has been observed that variables reflecting productivity or human capital, such as wage rates and education, are far from being the only factors driving gendered labor division phenomenon (Hersch and Stratton, 1994; Anxo *et al.*, 2002; Aronsson *et al.*, 2001; Rapoport and Sofer, 2005; Kalenkoski *et al.*, 2009). According to Rizavi and Sofer (2010), when a woman earns a higher wage than her partner, which is the case for around 20% of French households, there is no role reversal. In this case, the male partner's household work does increase as the woman's decreases. Women still do the major part of household work. Sevilla-Sanz *et al*. (2010) find some similar results using the Spanish Time Use Survey. In the US, the gender gap in non-market work is even higher when the woman earns more than her partner (Bertrand *et al*, 2013). Bittman *et al.* (2003) have found similar evidence in Australia. These observations are hardly reconcilable with standard household models assuming Pareto-optimality of household decision making. In addition, some studies question and reject the validity of the efficiency assumption in household models, especially on the production side of household decisions (Udry, 1996; Duflo and Udry, 2004).

Of course, a simple way to reconcile the empirical evidence described above and the Pareto-optimality in household production is to assume gender differences in the marginal product of labor in domestic production. Women could be more productive than men at home, with women's higher productivity more than compensating for wage differences when women's wages are higher. This is the assumption made by Becker (1981), based (possibly besides biological abilities) on the different training received by girls and boys through "gendered" toys and games. But today, in developed countries at least, though toys and games are still highly gendered, the skills necessary to perform domestic tasks are not really high any more. This follows the widespread use of market substitutes for domestic goods (buying clothes replaced sewing or knitting them), on the one hand, and household equipment in durable goods, on the other hand (it is not likely that women are better than men at pushing the "on" button of the vacuum cleaner!). Moreover, most fathers are now involved in the caring of children at all ages, including babies. Therefore, excluding the very limited time period of

pregnancy, and possibly breastfeeding, systematic productivity differences between genders again do not seem likely to occur in the raising of children either.

Recent household models move away from the productive efficiency mechanism. Persistent gender differences in preferences might also explain the aforementioned puzzle. Differences in preferences would occur if women, on average, "liked" performing domestic tasks more than men, due to habit, education or socialization. This incorporates a consumption component 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 the society could impose additional costs or gains to direct preferences when the individual deviates (or not) from these rules (Akerlof and Kranton, 2000). In this case, "gender roles" could impose a constraint leading households to deviate from the first best efficiency. Duflo and Udry (2004) show that in Côte d'Ivoire, the intra-household distribution of income and its uses are strongly constrained by social norms. Their empirical results lead to the rejection of the restrictions implied by Pareto-efficient household models. Non-cooperative household models including social norms have been developed by Lunberg and Pollak (1993), Carter and Katz (1997) and by Cudeville and Recoules (2013).

Inefficiencies in domestic production could also stem from the existence of gendered stereotypes within the society. Gendered stereotypes are the sets of beliefs grounded in the society that link personal attributes to the social categories women and men. For example, people could believe that women are more talented than men in domestic work and child care. A model of discrimination in the labor market involving stereotypes has been developed by Coate and Loury (1993), who assume stereotypes concerning differences in the productivity of black and white workers. A model of stereotypes in domestic production is developed in Sofer and Thibout (2011) and shows that stereotypes could explain the actual behavior of couples concerning the sharing of time spent in household production.

Until now, gender stereotypes have mainly been explored by psychologists. A significant experimental literature, published in psychology journals, analyzes the deleterious effect of the "stereotype threat" on individual performance. In a seminal study, Steele and Aronson

(1995) found that reminding people about a negative stereotype of their social group in a specific domain (e.g., African Americans and intellectual ability) impaired their subsequent performance on related tasks and thus confirmed the stereotyping they face. A large body of research has subsequently emerged to explore the effects of stereotype threats on performance: Schmader and Johns (2003), Cadinu *et al.* (2005) or Inzlicht and Ben-Zeev (2003), among others. What we explore here is merely the first approach: *gender stereotyping as a biased belief* rather than the "stereotype threat" which implies a negative impact on performances.

The objective of this paper is to test the efficiency of household time-allocation and to disentangle these different explanations by running an experiment with real couples. The practical inability to observe domestic production prevents the implementation of a convincing micro-econometric test. 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 in general. The experimental approach overcomes these difficulties by allowing a direct measure of individual domestic productivities and the perfect control of production technology. We can thus construct a real test of household productive efficiency.

A vast literature has now analyzed gender differences using an experimental framework (Croson and Gneezy, 2009). Yet, experiments with subjects that are real couples remain relatively rare. Ashraf (2009) tests efficiency of the public good contributions in couples from the Philippines. Cochard *et al.* (2009) test various social dilemmas with French couples, and compare couples' behavior to those of foreign couples. In our case, having true couples is necessary since we are specifically interested in observing the impact of stereotypes in the family context which should be characterized by a high level of communication and cooperation. Hence, when collecting evidence in this strong context, we would expect stereotypes to affect any kind of relationships in which the circulation of information is weak.

With the experiment described below, we want to explore if preferences, gender stereotypes and/or social norms affect couples in their household production decisions, and more precisely in the division of labor, so that their behavior may deviate from efficient behavior. Once productivity is measured, it becomes possible to examine if the sharing of household

time verifies the efficiency assumption, and, if not, do inefficiencies follow mainly from differences in preferences, or merely involve social norms or/and stereotype effects. This is the aim of the experimental design described below, in which we try to mimic the sharing of highly-gendered household tasks and compare it with the sharing of neutral tasks. Particular attention is paid to the information setting of the couples.

The experiment consists in asking couples to perform two day-to-day tasks in a fixed time, the free time saved by each partner over the allotted time being paid jointly to the couple, possibly at a different rate for the two partners. The accomplishment of the tasks thus comes as a fixed-cost, reducing the paid-time and consequently the couple's gain. We assigned three different treatments to this experimental design. The first treatment variable refers to the nature of the tasks involved: we consider one gendered setting (**G**), and one neutral setting (**N**). In the gender setting, the subjects have to perform "highly-gendered tasks". In the neutral setting, they have to perform tasks which are *a priori* neutral from gender. For "gendered tasks", we chose an activity involving clothes on the one hand (sorting out socks), and a repair activity on the other hand (attaching metal brackets). Indeed, European time use surveys show that domestic tasks implying 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. In Europe for example, Winquist (2004) compared 10 countries, and showed that the share of women's labor in the total household time devoted to laundry ranged from 80% in Sweden to 100% in Slovenia.[1] It is still worse for ironing, which is performed at 100% by women in Estonia, Slovenia, Hungary, Finland and Norway (in France the ratio was 94% by women against 6% by men). Conversely, "construction and repairs" is a masculine activity, as, on average, the sharing of time is around 85% for men to 15% for women in Europe (89% to 11% in France, the highest share for women being in the UK with a split of 80% for men and 20% for women). For the neutral tasks, we chose the tasks of filling and sealing envelopes, on the one hand, and writing down telephone numbers, on the other hand.

What we expected was that the sharing of time between partners would be less efficient in the gendered setting than in the neutral setting. In order to identify the potential role of alternative explanations in the deviations from efficiency, we took into account differences in preferences, social norm effects and stereotypes: beliefs about the partner's relative

productivities and preferences for the different tasks accomplished during the experiment were controlled for in questionnaires filled out during the experiment. One additional treatment was considered for that purpose: the information of the subjects about the performances of their partners in the tasks involved.

The remainder of the paper is organized as follows. Section 2 provides the theoretical benchmark for our experimental design. Section 3 details the experimental design and treatments. Section 4 presents the empirical strategy and the definition of the efficiency measure. In section 5, we discuss the experimental results. Section 6 concludes.

1. Theoretical Model and Inefficiency Measures

1.1 The theoretical model

We test the following theoretical model. Couples are expected to behave so as to maximize a pay-off function defined at the household level. This function is chosen to be linear in order to consider the simplest cognitive case for the volunteers. In the model, the level of domestic good production is fixed, so time inputs in domestic tasks are modelled as implying opportunity costs that need to be paid and shared within the couple. When choosing to share domestic tasks according to Ricardian comparative advantages, couples can minimize the opportunity cost of domestic production and maximize their gains in the labor market.

Formally, each couple, includes a man (m) and a woman (f), and is asked to perform two tasks, indexed by $i=\{1,2\}$, in a limited time, denoted T. For example, task 1 consists of sorting out 11 pairs of socks and task 2 consists of tightening 5 metal brackets in 7 minutes for the gendered setting. In the neutral setting, task 1 consists in filling and sealing 5 envelopes, and tasks 2 consists in writing down 5 telephone numbers. The total production of each domestic good is fixed experimentally and denoted D_i, with $\overline{D}_i > 0$. The production function of each good is linear with respect to the time inputs of both spouses, which are considered substitutable. The production processes of tasks 1 and 2 are independent. The time devoted by couple's members on task i is denoted t_f^i, for female time and t_m^i, for male time. The performance of each partner is assumed not to depend on time spent and quantity produced. Hence, men's and women's productivity in task i is fixed and denoted α_f^i and α_m^i.

Completing the two tasks entirely within the available time is the condition for the couple to obtain a positive monetary reward.

The pay-off function is defined at the couple level. It depends on the free time saved by each couple's member over the total available time T. Free-time choices can vary at different stages of the experiment, depending on individual wage rates which also vary.

The optimization problem of the couple can be written as:

$$(P)\begin{cases} \max \pi = w_f\left(T - t_f^1 - t_f^2\right) + w_m(T - t_m^1 - t_m^2) & (1) \\ \text{e. g. } \alpha_f^1 t_f^1 + \alpha_m^1 t_m^1 \geq \overline{D}_1 \qquad (\lambda \geq 0) & (2) \\ \alpha_f^2 t_f^2 + \alpha_m^2 t_m^2 \geq \overline{D}_2 \qquad (\mu \geq 0) & (3) \\ t_f^1, t_f^2, t_m^1, t_m^2 \geq 0 \end{cases}$$

where π is the pay-off function for the couple. The woman's and man's wage rates are respectively w_f and w_m, λ and μ are respectively the Lagrange multipliers associated with the constraints (2) and (3) of the couple's program.

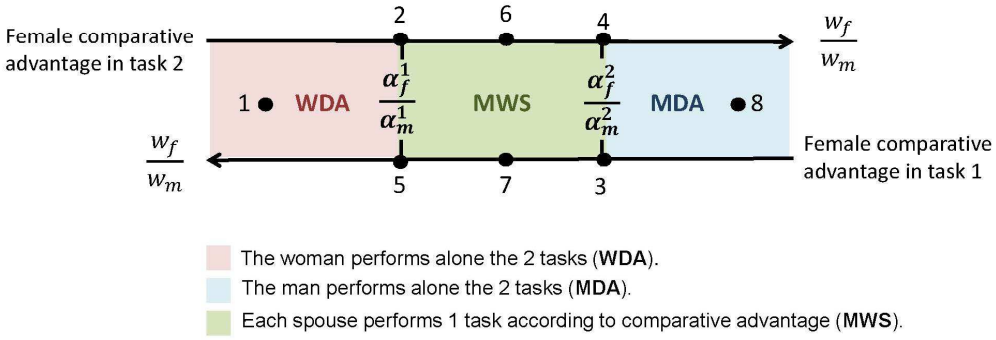
We restrict our attention to the cases where the man as well as the woman has enough time to perform alone both domestic tasks entirely, in the time available. So we focus only on solutions that satisfy the conditions: $\frac{\overline{D}_1}{\alpha_f^1} + \frac{\overline{D}_2}{\alpha_f^2} < T$ and $\frac{\overline{D}_1}{\alpha_m^1} + \frac{\overline{D}_2}{\alpha_m^2} < T$ (4).This specific case will be properly calibrated in the experimental design.

The first-order conditions for interior solutions are given by:

$$\begin{cases} t_f^1 \geq 0, \ \alpha_f^1\lambda - w_f \leq 0 \ and \ t_f^1\left(\alpha_f^1\lambda - w_f\right) = 0 \\ t_f^2 \geq 0, \ \alpha_f^2\mu - w_f \leq 0 \ and \ t_f^2\left(\alpha_f^2\mu - w_f\right) = 0 \\ t_m^1 \geq 0, \ \alpha_m^1\lambda - w_m \leq 0 \ and \ t_m^1(\alpha_m^1\lambda - w_m) = 0 \\ t_m^2 \geq 0, \ \alpha_m^2\lambda - w_m \leq 0 \ and \ t_m^2(\alpha_m^2\lambda - w_m) = 0 \end{cases}$$

These conditions define 8 different possible equilibria, depending on the values of the relative productivities of partners in the two tasks and their relative wage rate. These conditions are described in Figure 1. We can distinguish four different specialization regimes:

Figure 1. The Different Specialization Regimes of Spouses



- The first regime is the "Woman Does All" regime (**WDA**); point 1 in Figure 1 below. It occurs if the woman's relative wage rate is lower than her relative productivity in each task. In this case, the woman alone performs the two tasks, regardless of her comparative advantage, and the man saves all his available time.
- The second regime is the "Man Does All" regime (**MDA**); point 8 in Figure 1 below. It occurs if the woman's relative wage rate is higher than her relative productivity in each task. In this case, the man alone performs the two tasks and the woman saves all her available time.
- The third regime, the "man and woman share the domestic task" regime (**MWS**); points 6 and 7 in Figure 1 below. This occurs if the woman's relative wage rate lies between her relative productivity in each task. In this case, each partner alone performs the task for which he/she has a comparative advantage.
- The fourth, last regime (points 2, 3, 4 and 5 in Figure 1 below) corresponds to equilibrium solutions where the woman's relative wage rate is equal to her relative productivity in one of the two tasks. In this case, both partners share the corresponding task and the one who has the comparative advantage in the other task performs it alone.

The experiment consists in placing couples successively in regimes **MDA**, **WDA** and **MWS** in a random order, by setting appropriately the wage rates of the two partners, given their measured productivities in the two tasks. The fourth regime is excluded, as the corresponding equilibria 2 to 5 are too difficult to set experimentally, taking into account that, as can be expected, the measure of individual productivities implies some measurement errors.

1.2 Inefficiency measures

Inefficiency of task allocation within the couple is measured using the sole measure of output produced by one household member. This is because the total output is fixed for each task. For each task, the output produced by the man in any working session in couple can be derived from the output produced by the woman, as it is simply total output minus woman's output. The couple is considered efficient if it behaves according to the predictions of the model described in section 1

Formally, let us define the vector $\boldsymbol{q} = (q_1, q_2)$ where q_i denotes the quantity of good $i = (1,2)$ produced by the woman (the individual subscript is omitted for simplicity). This quantity is normalized by the total output required for this task (so it is divided by $\overline{D}_i$.). Let us denote the optimum $\boldsymbol{q}^*$. We recall that for the wage pattern **WDA**, $\boldsymbol{q}^* = (1,1)$; for the wage pattern **MDA**, $\boldsymbol{q}^* = (0,0)$; and for the wage pattern **MWS**, $\boldsymbol{q}^* = (1,0)$ if the woman has the comparative advantage in task 1 and $\boldsymbol{q}^* = (0,1)$ otherwise.

Our first measure of inefficiency, denoted d_1, is simply binary:

$$d_1 = \begin{cases} 0 \text{ if } \boldsymbol{q} = \boldsymbol{q}^* \\ 1 \text{ if } \boldsymbol{q} \neq \boldsymbol{q}^* \end{cases}$$

Our second measure of inefficiency, denoted d_2, is the squared Euclidian distance between the optimum and the choice made, it is:

$$d_2 = \|\boldsymbol{q} - \boldsymbol{q}^*\|^2$$

We then disentangle both distance measures to analyze if deviations from efficiency occur in the direction of the stereotype (the woman does more socks, task *i=1*, and/or less metal brackets, task *i=2*, than what the optimum would predict). This distance is denoted d_2^{exp}:

$$d_2^{exp} = \begin{cases} d_2 \text{ if } q_1 \geq q_1^* \; and \; q_2 \leq q_2^* \\ otherwise \end{cases}$$

The complementary case corresponds to the unexpected case where the deviation from the optimum goes in the opposite direction. In this case, we define d_2^{unexp}:

$$d_2^{unexp} = \begin{cases} d_2 \text{ if } q_1 \leq q_1^* \; and \; q_2 \geq q_2^* \\ otherwise \end{cases}$$

2. Experiment Design

2.1 General organization

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 new town outside central Paris), with a maximum of 4 couples per session, from March 2012 to January 2013. 81 couples, aged approximately 25 to 65 years old,[2] were recruited, essentially through a communication campaign using newspaper articles, radio and TV interviews. Couples must have lived in the same home together for a continuous period of at least one year (a condition checked in questionnaire II). Volunteers had been told that their couple could earn between €50 and €100 for sessions of approximately 2 hours, but they did not receive information about the purpose or the content of the experiment. To participate, couples had to register on a website created for the occasion.[3] They could obtain help from us thanks to a phone number that was available. Of course, we are aware of the existence of a participation bias in the recruitment treatment, since participation is voluntary and we did not observe the characteristics of the people who did not participate. Sessions were held on Saturdays, weekday afternoons or lunchtimes, to reduce the cost of participation. Still, we observed that in the sample of volunteers, couples belonged to a wide range of socioeconomic backgrounds, and different labor market situations (see Table IV).

Sessions lasted between 2 and 2½ hours. They were run by 5 to 6 experimentalists, and managed by one or two of the authors, with 4 couples participating in each session.[4] The experiment was a "pen and paper" exercise. Participants had to work on tasks that are not computerized. Time was measured by experimentalists who were also in charge of reporting production and ensuring that the payment rules of each game were perfectly understood by the participants. This kind of experiment takes time, it is more concrete and it is taken seriously by the participants. It guarantees a higher data quality than computerized experiments, except maybe for the data reporting which could generate mistakes but which was also strictly controlled by us. Another problematic issue related to the experiment design stems from the fact that participants could feel a lack of anonymity, due to the presence of one supervisory experimentalist per participant, who was therefore in close contact with

[2] The average age of husband and wife.

[3] www.expeco.fr

[4] We wish to thank all the experimentalists (colleagues, doctoral students and graduate students) who helped us in running this experiment as well as pilot sessions: Anil Alpman, Magalie Ayache, Marie Baguet, Sydia Batool, François Belot, Christophe Bournez, Roxane Bricet, Delphine Drouet, Thomas Erard, Gaelle Ferrant, Sekouba Fofana, Hélène Gace, Léontine Goldzahl, Fabian Gouret, Danielle Hadjiou, Lewis Hounpekevi, Ignacio Inoa, Stéphane Lavie, Nacim Laïmeche, Christophe Misner, Elodie Poncet, Juliette Rey, Thomas Richard, Omar Séné, Hervé Martial Takteu, Claire Thiboult, Stefania Velianoska, Yun Wu, Léa Yammine, and Carlos Torres Zuniga.

participants. Participants had to take their decisions freely without feeling any pressure from observation. This is why the team of experimentalists was trained by us to act professionally and to be as neutral as possible towards participants' decisions and performance, in order that they could feel free to take any decision they wanted, as well as to act in a natural way.

2.2 Detailed implementation of the experiment

The timeline of the experiment is given in Table I. At the beginning of each stage, instructions and payment rules were read aloud and projected on a screen.

Table I: Timeline of an Experimental Session

0)	**ALL PARTICIPANTS** - reception and general instructions	
1)	**INDIVIDUAL PART**	
	Training time for tasks 1 and 2	*Nature of the task varies across sessions (**G**ender/**N**eutral treatment)*
	Measurement of individual productivity tasks 1, 2	*Order randomized across participants.*
	Questionnaire I	
	Socio-demographic questionnaire	*Order randomized across sessions (stage 1, 3)*
2)	**COUPLE PART**	
	Hand out a paper with the 3 wage patterns (**WDA**, **MDA**, **MWS**) specific to the couple	
	Show individual productivities	*For sessions with informed treatment (**I**) only*
	Free discussion time for couple discussion and couple work 3 wage patterns	*Order of the 3 wage patterns (**WDA**, **MDA**, **MWS**) randomized across couples*
3)	**INDIVIDUAL PART**	
	Measurement of individual productivity, tasks 1, 2	*Order randomized across participants*
	Questionnaire III	
	Socio-demographic questionnaire	*Order randomized across sessions (stage 1. 3)*
4)	**PAYOUT** - Couple payout and good bye	

In stage 0, all the participants were informed together in a single room about the timeline and general payment rules. Participants were informed about the exchange rate of the experimental currency unit used in the experiment, which was **ECU35 = €1**. They were told that their decision was anonymous with anonymity guaranteed by personal identifiers that should be worn at all times. We also explained to them that the payment was given jointly, at the end, to the couple, with both members present. Participants were granted with a fixed payment of ECU700 = €20 for their participation. They were informed that additional money could be earned in the course of the experiment, depending on a current payment rule and their performance. In all stages, participants were told that the faster they accomplished the tasks, the more money their couple would earn. The experiment proceeded in three phases with a short break in between.

Stages 1 and 3 consisted in measuring individual productivities, *i.e.* time spent by each subject in performing each of the two tasks. In this individual work session, men and women were separated in different rooms. Each subject was assigned at a work station, spread out across the rooms and separated by screens. A workplace was made of U-shaped tables with 2 individuals working back to back. This configuration allowed one experimentalist to measure the performance of both participants at the same time. The subjects were asked to accomplish a fixed quantity of each of the two tasks in a minimum amount of time and were instructed that the faster they accomplished the tasks, the higher their income would be. They had first a training period to get familiar with the two tasks to be performed. Time measures were made on an individual basis by a timekeeper, discreetly posted at a distance. Individual productivity in each task was theoretically supposed to be fixed across games and time for one individual (α_j^i parameters of program (P)). The second measurement in the third stage aimed precisely at checking this point.

The maximum time allocated for individual production in these parts was set at *3 mn 30* for each task. This time was calibrated using pilot sessions to ensure that any participant could do the tasks in the time imparted. For task *i*, when the production D_i was reached, the experimentalist measured the time. The stopwatch was put on display on the work station so that the subjects could see their performance. Such discrete reporting was done to ensure there was no competition among the subjects. The difference between the maximum time and the time spent doing the task by the participant was paid at a rate of **ECU10 per second**. A short questionnaire aiming at eliciting beliefs and preferences was filled out during these phases

(see Questionnaires I and III in Appendices 1 and 2). The individual socio-demographic questionnaire (Questionnaire II)) was filled out during stages 1 or 3.

In stage 2, couples were asked to work jointly on the two same domestic tasks, in the same limited total time as in stage 1 *i.e.* with a maximum time which was the sum of the maximum times of each of the tasks. This treatment was repeated 3 times for each couple, with different wage patterns (**WDA**, **MDA**, and **MWS**). These wage patterns corresponded to the different regimes set out in the theoretical section. Using the individual measures of productivity for the first stage, the wage rates of each partner (*i.e.* valuation of individual free time) was individually fixed to ensure that the optimum fell within one specific regime. For example, to reach a **WDA** optimum, we could adequately fix the relative wage rate ratio w_f / w_m to be lower than the minimum relative productivities of women and men in each.

Couples were spread out across the rooms (separations between couples) and assigned a place within a specific U-shaped workplace. We first put the material on the tables and let the spouses place themselves naturally, wherever they felt like, inside this space. We trained the experimentalists to behave neutrally and very carefully in this stage, in order to avoid any bias coming from the observer. So, for example, the woman could choose initially with her partner to position herself closer to the material of one task or another: this would occur without any suggestion from the experimentalist. Before starting the work, we distributed a sheet of paper containing the three wage patterns that would apply to the three episodes of joint work.[5] In some cases (informed treatment), the paper also contained information about individual productivities. This information was also given aloud and explained precisely with video instructions. Before starting, the couple also had free discussion time to plan their work. This discussion time occurred at three intervals, before the application of each payment rule.

The couple was placed in a situation corresponding to the model set out in section 1. The partners had to complete the entire production of two tasks (D_1 and D_2) and choose the free time of each partner in order to maximize the pay-off (which is given in Program *(P)*, section 1). The time endowment T was calibrated so that the two tasks could be fully accomplished by one partner alone. This corresponded to the assumption we make in the theoretical model (see

[5] The computation of wage patterns requires some organization in a pen and paper experiment since we need to produce the sheet of paper containing the payment rules of each couple during the course of the experiment (a treatment that would take less than 1second in a computerized lab experiment but requires 10-15 careful minutes in ours).

inequalities (4) above). No couple of the sample failed in accomplishing the two tasks on time. Each individual could choose to take his/her free time at any moment. Free time was valued at the individual wage rates (denoted w_f and w_m). Decisions were jointly planned beforehand by the couple during the discussion time. We did not tell the participants that free time represented what could be labor market work. The rules of joint couple production of domestic goods were clearly explained to the participants. In short, we needed to ensure that the specific shape assumed in the theoretical part 1 corresponded to what we applied in the design (see more details in section 2.4). Concretely, each couple was given a limited time of 7 minutes to accomplish the same fixed quantity of the two same tasks as in the first stage. Couples were informed that they could choose to share the work as they wished. They could perform one task each, each of them being paid at his/her wage rate for the time remaining in the time allotted. They could instead choose to let one person complete the two tasks alone, the partner performing both tasks being paid at his/her wage rate for the time remaining, and the other at his/her wage rate for the whole available time. Or they could choose any intermediary solution, one of them performing both tasks, but with the participation of the other in at least one task or both. We reported time spent by each person in each task, as well as the amount of their production.

2.3 Treatments and predictions

In order to disentangle the potential effect of preferences, social norms and gendered stereotypes, we assigned two treatments to this experimental design according to different criteria.

Table II: Experimental Design 2x2

Treatments Gendered Tasks / Information	**Not informed (NI)**	**Informed (I)**	**Any**
Neutral (N)	21	20	**41**
Gendered (G)	21	19	**40**
Any	**42**	**39**	**81**

These treatments took place between sessions: neutral **(N)** vs gendered **(G)** tasks and informed **(I)** vs not informed **(NI)**. The number of observations in each case is summarized in Table II. The first criterion of the design is the nature of the tasks. The experimental sessions were divided in two types: gendered (**G**) and neutral (**N**). Gendered tasks are viewed as being 'naturally' attributed to a specific gender, whereas neutral tasks do not have this property. The treatment was chosen to be applied between subjects. It was not possible to implement a within treatment, as it would have involved the couples in the performing of four different tasks, which did not permit a reasonable duration for the sessions.

As we expected that stereotypes about gender roles change task allocation decision of a couple, our predictions were the following. When comparing the treatment group (couples performing gendered tasks) and the control group (couples performing gender neutral tasks), we expected to find more deviation from efficient choices in the treatment group in stage 2, with a higher number of men performing the "masculine task" (tightening metal brackets) and a higher number of women performing the "feminine task" (sorting out socks) than would result from an efficient division of labor. Of course, we expect that some deviations from efficiency would also occur in the neutral case, but the distance to efficiency would be higher in the stereotyped case.

In order to better understand what drives the previous result (if obtained), a second treatment (**I**,**NI**) controls the information flow within the couple. In all the sessions, in stage 1, it should be recalled that individual productivities are given orally and secretly by the timekeeper and visually (using the stopwatch), just after the measurement of performance by the experimentalist, in two different rooms for men and women. Then half of the couples (denoted group (**I**)) are informed about both partners' individual productivities measured in stage 1. Individual productivities (the time required by each member of the couple to perform each task) were given on the same paper sheet which contained the individual valuation of free time for the three wage patterns. In the (**NI**) group, the productivity information remains private, even within the couple. Of course, couple members have free discussion time when they can possibly share this private information.[6]

This treatment (**I**,**NI**) aims at disentangling the effects of a "social norm" from those of stereotypes: as set out above, stereotypes are the sets of beliefs grounded in society and which

[6] In practice, we never heard any of them ask their partner about his/her performance in tasks.

link a personal attribute to social categories. In our case, the beliefs would be that women are better at sorting socks than men, while men are believed to perform better than women at attaching metal brackets. If deviations from efficiency mainly come from the prevalence of stereotypes, information about the true performance of each partner should then modify these beliefs, at least partially, so as to decrease inefficiencies linked to stereotypes. Conversely, if the gendered division of labor reflects social norms rather than stereotypes, there will be a cost in deviating from the social norm, whatever the true performance in completing task. What is important then for subjects is "doing gender" (Sevilla-Sanz *et al.*, 2010). If people consider, to some extent, that it is the "role" of women to deal with clothes, while it is that of men to deal with metal brackets, they will choose to share work in that way, even if this choice has a cost in terms of efficiency. Information plays no role in this case. Hence, better information about true performance should not significantly modify the decision treatment if social norms play a prominent role in the gendered division of labor.

Finally, to complete this design, we make the participants fill out an individual socio-demographic questionnaire, as well as two questionnaires about their beliefs and preferences (Questionnaires 1 and 2; the timeline is given in Table I above). Anonymity of the answers was guaranteed, even within couples.

2.4 Tasks and production rules

The tasks were chosen according to the following requirements. They should be simple enough not to leave room for variation in quality. Performance of participants should be different but rather stable in time, so we could rule out learning or tiredness effects. Of course, the tasked had to be doable manually in a class room, and not too costly. The production had to be easy to count and divisible.

Gendered Tasks

The female-oriented gendered task consisted in sorting out and forming balls of folded socks with a matched pairs, from a basket containing 11 potential pairs and 10 unmatched socks. These socks were of different colors, sizes and with different patterns. They were mixed up in the basket. The basket was changed between rounds. Regarding the production rules, we only allowed the participants to take the socks out of the basket once the ball of folded pairs was

formed, not before. The production level to be reached was 11 pairs of socks, within a maximum time of 3½ minutes.

The male-oriented gendered task consisted in attaching 2 metal brackets with 2 nuts and 2 screw bolts, five times. Each shape should form a 'T', to be tightened without any play. The nuts and bolts could be introduced in different directions. The material was presented in a basket with more nuts and screw bolts than necessary, but the exact amount of metal brackets. The production level to be reached was 5 'T' shapes, with a maximum time of 3½ minutes.

Neutral Tasks

To perform the first neutral task, participants needed to stuff one envelope with two sheets of paper of different colors, appropriately folded in three, and then close the envelope. The material put on the table contained exactly the right amount of envelopes and 2 piles of colored sheets of paper. The required production level was 5 envelopes, in a maximum time of 3½ minutes.

The second neutral task was to search and find 5 telephone numbers in a list of 50 names matched with phone numbers, and write them down on five pieces of paper, each with a name of the list
on it. Material for this task included a list of two pages containing names matched with phone numbers, pens, and five nominative forms. The name list was different in each round. The required production level was 5 phone numbers, in a maximum time of 3½ minutes.

Joint production rules for couples

In the gendered setting, the required level of joint production within a couple was 5 'T' shapes and 11 pairs of socks in 7 minutes maximum (for both). In the neutral setting, the couple had to produce 5 envelopes and 5 phone numbers in 7 minutes maximum (for both). The rules of joint production for a couple were precisely described to the participants with lots of illustrative examples. In order to avoid complementarity between spouses' time, we forbid the participants to help their partner in the production process of one good. For example, one partner was not allowed to indicate to his/her partner if he/she saw the matched sock. In the production of metal bracket 'Ts', one partner was not allowed to give the bolts to the other

and so on. We also insisted on the fact that they could switch tasks if they wanted at any time, they could also work together in the same basket at the same time without helping each other. In each round, free time should be taken once and for all by any of the partners by declaring "free time" to the experimentalist.

To give an example, assume that the different time values for each partner (F and M), i.e. their productivity in performing tasks are the following, shown in lines 1 and 2 of Table III below:

Table III: Example

Time spent at performing (minutes.sec)	**Socks**	**Brackets**
F	**3.30**	**2.15**
M	**2.30**	**3.30**
Wages in each regime (ecus/s)	w_f	w_m
MWS	**300**	**300**
MDA	**404**	**196**
WDA	**106**	**494**
	Stereotype in each regime	**Efficiency in each regime**
Total gains (euros)	**60**	**72.5**
[ecus]	**[2100]**	**[2537,5]**

Then, the wages computed in each regime would be those appearing in lines 3, 4 and 5 of Table III. Total gains of the couple would depend on the way they share tasks in the three regimes. Here, the woman has a comparative advantage (even an absolute advantage) in brackets as compared to socks relative to her partner: efficiency would require that she tightens the brackets and he sorts out socks in the MWS regime. This is the way of saving most free time for both, paid at the same wage rate. Furthermore, according to the design of the experiment, taking into account the wage ratios of their free time, he should do both tasks in the MDA regime while she does nothing, and conversely she should do all while he does nothing in the WDA regime. If the couple chooses efficiency in the three regimes, its total

gains[7] would reach a maximum of 72.50 euros. If, conversely, they systematically share tasks according to gender roles, i.e. he tightens the brackets while she sorts the socks in the three regimes, the couple's gains would be 60 euros. Hence, efficiency raises the couple's gains by more than 20% in this example.

3. Results

Tables IVa and IVb below give the main characteristics of the sample. The couples were on average 40 years old. The age difference between men and women was on average two years, which is usual for French population data.

Table IV: Descriptive Statistics

Table IVa Individuals Characteristics

	Women	**Men**
Age	39.58	41.54
	(12.06)	(12.18)
Education Level (years of education)	16.56	18.74
	(13.57)	(12.53)
Employed	0.8025	0.9136
	(0.3989)	(0.2815)
Individual Account	0.8395	0.7531
	(0.3678)	(0.4321)

Table IVb Household Characteristics

	Cergy	**Paris**
Average Age	40.26	40.92
	(12.26)	(11.49)
No child	0.4773	0.5405
	(0.5004)	(0.4995)
Married	0.4204	0.3243
	(0.4945)	(0.4692)
Civil Union ("Pacs")	0.1363	0.2162
	(0.3438)	(0.4126)
Not married, nor "Pacsed"	0.4432	0.4595
	(0.4977)	(0.4995)
Couple duration (years)	11.87	13.16
	(11.47)	(10.96)
Joint Account	0.6364	0.6622
	(0.4820)	(0.4740)

[7] Total gains include gains from individual sessions of the experiment. We assume in the example that the time spent by each partner in each task remains the same –shown in Table III above- in all sessions. More productive couples could earn more than 100 euros.

The couples were stable: they had lived together for more than 10 years on average. They had relatively high employment rates and education levels. Selection according to these observable characteristics was comparable in Cergy and in Paris. In the Paris sample of volunteers, couples were less often married and had children less often too. They were also more often employed, which is also what was to be expected from the population data relating to inner Paris, compared to its suburbs.

Result 1: Having to perform gender stereotyped tasks reduces the likeliness of choosing an efficient division of labor.

Table V below shows a comparison of efficiency measures using the discrete and metric distance variables (see above, section 1.2) for the three rounds of stage 2, corresponding respectively to the three wage patterns **MWS**, **WDA** and **MDA**. We observed that in the MWS case, where efficiency required that each partner specialized entirely in the task corresponding to her/his comparative advantage, couples achieved efficiency in more than 73% of cases for neutral tasks, while they achieved efficiency in only 60% of cases for gendered tasks. This difference between their performances in neutral tasks and in gendered tasks is shown to be systematic, with the difference in performances varying to some extent, depending on the type of equilibrium. It should be noted that for the **MDA** and **WDA** wage patterns, couples were more often inefficient than efficient, whatever the type of task, while the opposite occurred for specialization (the **MWS** treatment). This could potentially be due to zero aversion (aversion to do nothing while the partner does everything).[8] However, for the three types of equilibrium, and according to what we expected, the percentage of efficient choices was higher for neutral tasks than for gendered tasks.

We test whether the groups submitted to the gendered treatment had a significantly different efficiency behavior compared to the groups submitted to the neutral treatment. We model the probability of being efficient using the data pooled over the 3 rounds (243 observations). We use the Mantel-Haenszel (1959) test to check whether the effect of the gendered treatment was heterogeneous or could be considered as homogenous over the MWS, MDA and WDA decisions. 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, MDA and

[8] Note, however, that as participants knew the three wage settings from the beginning, they also knew that they would have the opportunity to reverse roles in another round: this design was used precisely to avoid equity problems.

WDA). The test passes at the 5% level with a p-value of 0.6616, so we chose to implement the tests of the impact of the gendered treatment pooling the 3 rounds. Regarding the discrete measure of efficiency, a Mann-Whitney-Wilcoxon (Mann and Whitney, 1947; Wilcoxon, 1945) non-parametric test is used to check whether the efficiency rate is significantly different for groups treated with the gendered as compared to the neutral group. This test can be applied on unmatched data in which individuals of the groups can be supposed to have behaved independently of each other, which was indeed the case here. The p-value equals 0.0281 and leads to the rejection of equality of means, at the 5% level. Regarding the distance to efficiency measure (last column on the right), the Mann-Whitney-Wilcoxon test statistics leads to a p-vale equals to 0.0810. So we reject the equality of means at the 10% level.

Table V: Impact of the Gendered Treatment on Couple Efficiency

	Percentage of Efficient Couples (measure $1 - d_1$)		**Distance from optimum** (measure d_2)	
Treatment	**Neutral**	**Gendered**	**Neutral**	**Gendered**
MWS	73.2%	60.0%	0.493	0.532
(n=81)			*(0.867)*	*(0.809)*
MDA	20.0%	15.0%	0.745	0.818
(n=81)			*(0.451)*	*(0.449)*
WDA	48.8%	25.6%	0.458	0.672
(n=81)			*(0.512)*	*(0.273)*
All	**47.5%**	**33.6%**	**0.564**	**0.674**
(n=243)			*(0.646)*	*(0.607)*
Mann Whitney test (p-value)	***0.0281 *****		***0.0810 ****	

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

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

Table VI below illustrates this result. In Table VI, the distance to optimum has been decomposed in the following way for the gendered treatment: a deviation from the optimum is characterized as "expected" (going in the expected direction), if the deviation is made toward the stereotyped value (the woman does more socks and/or the man does more metal brackets

than what the optimum would dictate), while it is characterized as "unexpected" (going in the unexpected direction) if the deviation from optimum goes away from the stereotyped value (see above, section 1.2).

It can be seen in Table VI below that, as could be expected, couples do deviate more towards the stereotype than in the opposite direction, whatever the distance measure used here (discrete or metric). The tests show a highly significant difference between the two directions of deviation: clearly, for couples, the stereotyped behavior has power to draw them away from efficient behavior.

Table VI: Distance from Efficiency, Modified Measure

For Gendered treatment (n=119)	To optimum	Expected direction	Unexpected direction	Test expected different from unexpected*	Test expected distance > unexpected
Distance d_2	0.6741	0.4588	0.2153	Pvalue	Pvalue
(standard error)	(0.6072)	(0.5838)	(0.4396)	0.0022 ***	0.9989

* Mann-Whitney-Wilcoxon sign-rank non-parametric test on unmatched data.

Result 3: Gender differences in preferences for tasks cannot explain the efficiency difference between the gendered and the neutral treatment (Result 1)

Preferences for tasks were collected using an individual anonymous questionnaire filled in by the participants during the experiment. Table VII below shows that, on average, in the neutral sessions, women prefer envelopes, while men prefer phone numbers; and in the gendered sessions women prefer socks, while men prefer metal brackets.

Table VII: Preferred Task

Neutral Treatment	Phone No	Envelopes
Women (n=41)	48.8 %	51.2 %
Men (n=39)	87.2 %	12.8 %
Gendered Treatment	**Socks**	**Squares**
Women (n=40)	72.6 %	27.4 %
Men (n=40)	42.5 %	57.5 %

Surprisingly, if we look at the differences in preferences, we can see that the highest gender difference preference for tasks appears for phone numbers and envelopes: there is nearly a 75-percentage point difference in favor of the phone number preference by men, compared to envelopes. Women do prefer socks over squares, but the difference is much lower: 45 percentage points only. The difference for men between socks and metal brackets is very small: 5 percentage points in favor of metal brackets, and the difference for women between phone numbers and envelopes is about 2 percentage points only. It can easily be concluded from these figures that the deviation from efficiency towards the stereotyped value is not driven by preferences; otherwise the deviation from optimum values would have been observed to be larger for neutral tasks than for stereotyped tasks, which is not the case.

Result 4: Beliefs about gender differences in performance are strongly stereotyped, compared to real gender performance differences

Let us now turn to relative performances and to beliefs about them. Table VIII below shows that, on average, women perform better than men in all tasks (less time, hence better performance). We recall that these measures were carried out during the individual work session. The gender difference in performance appears lowest for phone numbers, and highest for socks. We computed comparative advantage from individual performance measures,. Table IX shows the observed comparative advantage for women. Recall that a woman will have a comparative advantage in metal brackets if her relative productivity (inverse of the ratio of time spent) in attaching metal brackets compared to matching socks is better than the relative performance of her male partner (see Figure 1). In this case, her partner will have a comparative advantage in socks.

TABLE VIII: Gender and Observed Performances for Tasks

Treatment	Tasks	Male/Female Time
Neutral	Phone	1.11
	Envelopes	1.18
Gendered	Socks	1.25
	Squares	1.10

We observe from Table IX that women more often have a comparative advantage in stuffing envelopes (in nearly 60% of the observations) as compared to selecting phone numbers, and an even greater comparative advantage (60% of the observations) in matching socks as compared to attaching metal brackets. Remember, however that wages were set in order to control for these differences.

Table IX: Who has the Comparative Advantage?

Treatment	Tasks	Woman has the comparative advantage
Neutral (n=41)	Phone	41.46 %
	Envelops	58.54 %
Gendered (n=40)	Socks	60.00 %
	Squares	40.00 %

Beliefs about performance were collected via a questionnaire (Questionnaire 1, see Appendix A1). The results are shown in Table X.

Table X: Beliefs about Relative Performance of Women Compared to their Partner

Treatment	Task	Beliefs According to	Woman faster than partner	Equal Performance	Woman slower than partner
Neutral (n=41)	**Phone**	Women	43.90%	21.95%	34.15%
		Men	30.77%	25.64%	43.59%
	Envelopes	Women	48.78%	24.39%	26.83%
		Men	43.59%	33.33%	23.08%
Gendered (n=40)	**Socks**	Women	56.41%	15.38%	28.21%
		Men	47.50%	37.50%	15.00%
	Metal brackets	Women	17.50%	17.50%	65.00%
		Men	25.00%	22.50%	52.50%

We observe that, in line with gendered stereotypes, 56% of women (47.5% of men) believe that the woman in their couple is faster than her partner in the "socks" task (65% of women and 55% of men believe that she is at least as fast as him), and that, symmetrically, 65% of women and 52% of men believe that men will be faster in the task of attaching metal brackets. Regarding neutral tasks, it can be observed that beliefs about performance are gender differentiated, but much less so than for stereotyped tasks: women do not make any significant difference between phone and envelopes, either concerning their own performance or that of their partner. More men, on average, tend to believe that themselves will be better than their partner at phone numbers (44% of them, while only 34% of women have the same belief concerning the relative performance of their male partner). Women's and men's beliefs about envelopes are more similar and in favor of women (only 27% of women and 23% of men believe that the woman in their couple will be slower than her male partner). Note that, in general, looking at neutral tasks, each gender seems to have a better opinion of its own performances.

Table XI: Observed Relative Performance of Women Compared to their Partner

Treatment	Task	Woman faster than partner	Equal Performance (+/-5%)	Woman slower than partner
Neutral (n=41)	**Phone**	51.22%	9.76%	39.02%
	Envelopes	56.10%	14.63%	29.27%
Gendered (n=40)	**Socks**	65.00%	12.50%	22.50%
	Metal brackets	47.50%	17.50%	35.00%

Table XI represents the relative observed performance in the couple. It can be compared with the beliefs of Table X. The threshold for equality has been set at +/-5%.[9] Interestingly enough, it can be seen that women's performances are always underestimated, by men still more in general than by the women themselves. Note however that the difference between beliefs and true values is not so high in the "feminine task", *i.e.* socks. In more than 75% of the observations, women are observed to be at least as fast than men in performing this task. It can be compared with beliefs, where about the same percentage of the participants share the

[9] The results are robust to another choice of the threshold (+/-1% or +/-10%)

same opinion. The difference between observed performance and beliefs occurs for the "masculine task". In the case of metal brackets, the difference between true figures and beliefs is very high: 47.5% of women appear clearly faster than their spouses in this task (respectively 65% at least as fast), whereas only 22% (resp. 40%) of the participants share this opinion. It should be noted that, interestingly, men seem to be less influenced by the stereotype than women. Differences between beliefs and performances for the masculine task are higher for females than for males: women seem to underestimate their true performance in the masculine task more. Regarding the neutral tasks, performance and beliefs also differ, women's performances being always underestimated, but the difference is less marked than for the gendered task. Men often believe that they will perform as well as their partner though, in fact, she performs better. This is especially true for phone numbers, (30% of men only think that their partner will be faster though the true figure is 51%), but, overall, even in that case, their feeling is not so bad as "phone" is the task for which the percentage of women being slower than their partner is the highest, very close to men's beliefs. It may be concluded that couples share false beliefs about women's and men's relative performances in gendered tasks: the performance of women in the "masculine task" is clearly underestimated.

Result 5: Giving information about performance within the couple reduces the stereotype effect.

Table XII: Effect of Informational Treatment

Table XIIa: On Efficiency Rates

Treatment	**Number of couples**	**Efficient**
Informed (n=126)	**42**	40.5 %
Not Informed (n=117)	**39**	40.8 %

Table XIIb: On Distance to Optimum, Gendered Tasks

Treatment	**Distance Measure d_2**		
Informed **(**n=56**)**	0.6019 *(0.5698)*	Expected	0.3578 *(0.5173)*
		Unexpected	0.2441 (0.4847)
Not Informed (n=63)	0.7382 *(0.6362)*	Expected	0.5486 *(0.6275)*
		Unexpected	0.1896 *(0.3973)*

Finally, Table XII above shows the difference between the deviation from the optimum obtained in the informed and the non-informed treatment. As can be seen in the table, the informational treatment, which consists in giving the productivity information of partners in the couple before the couple task allocation game, has generally a modest effect on the number of efficient outcomes. Conversely and interestingly enough, information tends to reduce the measured distance to the optimum, which is especially true for the expected distance. A formal statistical test does allow us to confirm this result fully: the deviation from efficiency towards the stereotype increases in the absence of information (expected deviations), whereas it is reduced when the information is present. This result is another piece of evidence supporting our interpretation that stereotypes could be one of the main causes of inefficiencies in couples' choices of time allocation.

Conclusion

The design of the experiment aimed at showing that couples tend to divide labor in an inefficient way, especially when they are performing "gendered tasks". 81 couples participated in an experiment which was run partly in Paris and partly in the suburban new town of Cergy-Pontoise. After having measured individual productivity for separate tasks, we let the couples choose together how to carry out jointly both tasks, in a limited time. Theory predicts that couples should allocate their time according to their comparative advantage. Analyzing the results, we prove that gendered tasks induce couples to deviate from efficiency most of the time. Couples significantly deviate more, and more often when they perform gendered tasks as compared to neutral tasks. They also deviate more often in the expected direction, *i.e.* towards the stereotyped value, where the woman does more pairing of socks and the man does more tightening of metal brackets.

Our results also provide further explanations of the results. Three main causes for such deviations are supplied by theory: *i.e.* 1/ preferences directly entering the production function; 2/ social norms (there is a cost when deviating from "gender roles"); and 3/ stereotypes (false beliefs about respective productivities in "gendered tasks" of women and men). The results of our experiment make it possible to exclude the first cause. Moreover, though our results do not provide direct evidence against the role of social norms (gender roles) as sole explanation of inefficiency, they do so indirectly, as we show that information matters. The results indeed show that false beliefs based on stereotypes about men's and women's relative productivities

in tasks, which can be characterized as “masculine or “feminine”, were significant in leading to inefficiencies in the allocation of tasks among the couples participating in our experiment.

Appendix A1: Questionnaire I

N°Anonymat : _____

Date : _____

Session : _____

Individual and Anonymous Questionnaire I

1) If you could choose to realize only one of the two tasks proposed, which of them would you choose:
   - □ the task «phone number»/ «socks»
   - □ the task « envelopes »/ « squares»

2) Do you think that you will be:
   - □ faster on the task «phone number»/ «socks»
   - □ faster on the task « envelopes »/ « squares»
   - □ equally fast on both tasks

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

   on the task «phone number»/ «socks»:
   - □ faster
   - □ less fast
   - □ equally fast

   on the task «envelopes»/ « squares»:
   - □ faster
   - □ less fast
   - □ equally fast

Appendix A2: Questionnaire 3

N° : _____
Date : ____
session : __

Final Individual Questionnaire

1) Did you find each realized task: ?

Phone /: 1 □ Very easy
Socks 2 □ Rather easy
3 □ Rather difficult
4 □ Very difficult

Envelopes/: 1 □ Very easy
Squares 2 □ Rather easy
3 □ Rather difficult
4 □ Very difficult

2) Did you find each realized task: ?

Phone /: 1 □ Very pleasant
Socks 2 □ Rather pleasant
3 □ Rather unpleasant
4 □ Very unpleasant

Envelopes/: 1 □ Very pleasant
Squares 2 □ Rather pleasant
3 □ Rather unpleasant
4 □ Very unpleasant

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

……

……

……

……

Thank you for your participation!

References

Akerlof G. A. and R. E. Kranton (2000), *Economics and Identity*, Princeton University Press

Aguiar, M. and E. Hurst (2007), "Measuring Trends in Leisure: The Allocation of Time over Five Decades", *The Quarterly Journal of Economics*, MIT Press, vol. 122(3), pages 969-1006, 08.

Andreoni, J. and L. Vesterlund (2001), "Which is the Fair Sex: Gender Differences in Altruism", *Quarterly Journal of Economics*, CXVI, 293–312.

Anxo, D., Flood, D. 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*, 352-35: 127–150.

Aronsson, T., Daunfeldt, S.-O., and M. Wikström (2001), "Estimating intrahousehold allocation in a collective model with household production", *Journal of Population Economics*, 14(4): 569-584.

Ashraf, N. (2009), "Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines", *American Economic Review*, 99(4): 1245-1277.

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

Bertrand, M., E. Kamenica and J. Pan (2013), "Gender identity and relative income within households", *Working Paper*, University of Chicago

Bittman, M., P. England, L. Sayer, N. Folbre and G. Matheson (2003), "When does gender trump money ? Bargaining and time in household work", *American Journal of Sociology*, 109: 186-214.

Blau, F. and L. Kahn (2006), "The U.S. gender pay gap in the 1990s: slowing convergence", *IZA Discussion Paper*, 2176

Cadinu, M., A. Maass, A. Rosabianca and J. Kiesner (2005), "Why do women underperform under stereotype threat? Evidence for the role of negative thinking", *Psychological Science*, 16(7), 572-578.

Carter M and E. Katz (1997), "Separate Spheres and the Conjugal Contract: Understanding the Impact of Gender-Biased Development". *In: Haddad L, Hoddinott J, Alderman H (Eds). Intrahousehold Resource Allocation in Developing Countries: Methods, Models and Policies.*

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

Cochard, F., H. Couprie and A. Hopfensitz, (2009), "Do spouses cooperate ? If not: why?", *THEMA Working paper*, 2009-10

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

Cudeville E. and M. Recoules (2014), "Household Behavior and Social Norms: A Conjugal Contract Model", forthcoming in *the Annals of Economics and Statistics Special Volume: Economics of Gender.*

Duflo E. and C. Udry (2004). "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices", *NBER Working Papers*, 10498, National Bureau of Economic Research.

Dumontier, F., D. Guillemot et D. Méda (2002), "L'évolution des temps sociaux au travers des enquêtes Emploi du temps", *Économie et Statistique*, 352-353, 3-13.

Goldschmidt-Clermont, L. and E. Pagnossin-Aligisakis (1995), "Measures of unrecorded economic activities in fourteen countries" UNDP, Background Papers for the *Human Development Report*, New York: Oxford University Press: 105-155.

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

Inzlicht M. and T. Ben-Zeev (2003), "Do high-achieving female students underperform in private? The implications of threatening environments on intellectual treatmenting", *Journal of educational Psychology*, 95, 796-805.

Kalenkoski, C.M., Ribar, D.C. 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), 399-419.

Mann, H. B. and D. R. Whitney (1947), "On a test whether one of two random variables is stochastically larger than the other", *Annals of Mathematical Statistics*, 18, 50-60.

Mantel N. and W. Haenszel (1959), "Statistical aspects of the analysis of data from retrospective studies of disease", *J Natl. Cancer Inst*., 22: 719-748.

Miranda, V. (2011), "Cooking, Caring and Volunteering: Unpaid Work around the World", *OECD Social, Employment and Migration Working Papers 116*, OECD Publishing.

Rapoport, B. and C. Sofer (2005), "Pure Production Factors and the Sharing Rule: Estimating Collective Models with Household Production", *Document de travail, MSE*, Université Paris 1-Panthéon-Sorbonne.

Rizavi S.S. and C. Sofer (2010), "Household division of labor: Is there any escape from traditional gender roles?", *Working paper, Université Paris 1*, Cahierts de la MSE 2010.09.

Schmader T. and M. Johns (2003), "Converging evidence that stereotype threat reduces working memory capacity", *Journal of Personality and Social Psychology*, 38, 440-452.

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

Sofer, C. and C. Thibout (2011), "stereotypes upon abilities in domestic production and household behaviour", *CES Working Paper*, 2011.75

Steele and Aronson (1995), "Stereotype Threat and the intellectual performance of African Americans", *Journal of Personality and Social Psychology*, 69, 797-811.

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

Wilcoxon, F. (1945) "Individual comparisons by ranking methods", *Biometrics*, 1: 80-83.

Winqvist, K. (2004), "How Europeans spend their time. Everyday life of women and men", *European Comission Report, Pocketbook Edition, Luxembourg*.