B Results from Experiment 1

B.1 Norms regarding gender roles

Tables 2 and 3 report the distributions of the social appropriateness ratings of different alternative decisions under two the framing scenarios. We construct the mean ratings by converting responses into numerical scores as in ?. "Very socially inappropriate" is -1, "somewhat socially inappropriate" is -1/3, "somewhat socially appropriate" is 1/3, and "very socially appropriate" is 1.

In total, there were 38 subjects in the "Stay framing" treatment. They were incentivized to provide a rating regarding the social appropriateness of three possible actions the couple could take in Experiment 2, regarding who should keep working for money. 26 were in the "Quit" framing condition, where they were provided questions describing how the couple had to choose who would quit their job (do the task without payment in Experiment 2). In addition, they also had to guess what would be the most frequent actual choice in Experiment 2.

In the "Stay framing" scenario, couples choosing husbands to keep working is considered as the most socially appropriate option. The modal response to "husbands keep working" is that it is somewhat socially appropriate. In contrast, "wives keep working" is rated as the least socially appropriate, and the modal response is "somewhat socially inappropriate".

Similarly, when we ask subjects to rate different options in the corresponding experimental scenario, but with the terms boyfriends and girlfriends rather than husbands and wives, it is also considered more socially appropriate for boyfriends rather than girlfriends to keep working for payment. In both the "husbands and wives" framing and "boyfriends and girlfriends" framing, flipping a coin to decide who continues to work is rated as the most socially appropriate.

In the "Quit framing" scenario, husbands staying at home to take care of kids is considered socially inappropriate. Wives staying at home is considered as appropriate. Consistently, boyfriends working without payment and depending on their partner financially is viewed as socially inappropriate, while girlfriends working without payment is rated as socially appropriate. Flipping a coin to decide is judged to be the most socially appropriate.

We show the results regarding the empirical or descriptive norms about gender roles in Table 4. The table summaries counts of subjects' guesses regarding which option is the most frequently or the least frequently chosen by couples under each scenario. Subjects in the "Stay framing" scenario treatment only provide guesses concerning who works for payment, while subjects in the "Quit framing" scenario only answer questions regarding who works without payment.

The majority guesses that boyfriends working for money would be the most frequent choice, while girlfriends working for money would be chosen by the fewest couples.

However, there is no apparent empirical norm governing who works without money.

Altornativos	"Stay framing" scenario (N=38)						
Alternatives	Mean		-	+	+ +		
Husbands keep working	0.51	0.00%	7.89%	57.89%	34.21%		
Wives keep working	0.02	2.63%	47.37%	44.74%	5.26%		
Flipping a coin to decide	0.25	18.42%	18.42%	21.05%	42.11%		
Boyfriends keep working for payment	0.46	0.00%	10.53%	60.53%	28.95%		
Girlfriends keep working for payment	0.25	2.63%	21.05%	63.16%	13.16%		
Flipping a coin to decide	0.67	5.26%	5.26%	23.68%	65.79%		

Table 2: Elicited norms in the "Stay framing" scenario

Note: Following Krupka and Weber (2013), responses are: "very socially inappropriate" (--), "somewhat socially inappropriate" (-), "somewhat socially appropriate" (+), "very socially appropriate" (+). To construct the mean ratings, we converted responses into numerical scores ("very socially inappropriate" = -1, "somewhat socially inappropriate" = -1/3, "somewhat socially appropriate" = 1/3, "very socially appropriate" = 1)

Table 3:	Elicited	norms	in	the	"Quit	framing"	scenario	

Altornativos	"Quit framing" scenario $(N=26)$						
Alternatives	Mean		-	+	+ $+$		
Husbands stays at home	-0.21	7.69%	65.38%	26.92%	0.00%		
Wives stay at home	0.31	0.00%	19.23%	65.38%	15.38%		
Flipping a coin to decide	0.41	15.38%	3.85%	34.62%	46.15%		
Boyfriends work without payment	-0.03	3.85%	46.15%	50.00%	0.00%		
Girlfriends work without payment	0.05	3.85%	34.62%	61.54%	0.00%		
Flipping a coin to decide	0.64	3.85%	3.85%	34.62%	57.69%		

Note: Following Krupka and Weber (2013), responses are: "very socially inappropriate" (--), "somewhat socially inappropriate" (-), "somewhat socially appropriate" (+), "very socially appropriate" (+). To construct the mean ratings, we converted responses into numerical scores ("very socially inappropriate" = -1, "somewhat socially inappropriate" = -1/3, "somewhat socially appropriate" = 1/3, "very socially appropriate" = 1)

T 11 4	T • • 1		a 1 · · ·		1 /	1 1	, 1	•
Inhlo /It	Empirical	norme	Subiocte'	MILOGGOG	about	othorg	roportod	VIOWE
Lanc 4.	Empirica	norms.	BUDIEUIS	e ucooco	about	OULTER	TEDOLIEU	WS
	T			0			· · · · · · · ·	

	Stay-fi	aming	Quit-framing		
	(Choose one to	work for money)	(Choose one to work without money)		
	Guess most couples' choices	Guess least couples' choices	Guess most couples' choices	Guess least couples' choices	
Boyfriend	18	2	10	9	
Girlfriend	1	12	7	9	
Computer decides	5	10	9	8	

B.2 Gender differences in the preference/cost of doing the task

Figure 1 illustrates the maximal amount of money, in terms of the experimental currency ECU, that subjects are willing to pay to avoid the task. This is a measure of the disutility of doing the task. The graph shows that most people do not like performing the task, as most of them are willing to pay at least 1 ECU to avoid the task. On average, females are willing to pay 7.7 ECU and males are willing to pay 5.8 ECU to avoid the task. However, this gender difference is not statistically significant (p = 0.2555 for the two-sided t-test, p = 0.2387 for the Wilcoxon rank-sum test, p = 0.159 for the two-sample Kolmogorov –Smirnov test). Most people are not willing to pay more than 15 ECU to avoid the task.

We also check the correlation between individuals' performance and their disutility of doing the task. We are interested in knowing whether better performers have a lower disutility for the task. In general, there does not exist a significant relationship between performance in the task and disutility (the Pearson's correlation coefficient is 0.0272, with p = 0.8312). This relationship remains insignificant for both females and males considered on their own. If we analyse the correlation for women and for men separately, the Pearson's correlation coefficient is 0.2868 for females (p = 0.1115) and -0.1715 for males (p = 0.3480). These results imply that both genders exhibit disutility towards doing the task, while there is no essential difference in this distaste between genders.

Figure 1: Histogram of the maximal ECU participants are willing to pay to avoid the task, by gender



B.3 Gender differences in preference for being the money-earner

As we show in Figure 2 males have a somewhat greater willingness to work as the moneyearner for the group on average. However, the difference is not statistically significant. Using a non-parametric Wilcoxon rank-sum test yields p = 0.3075, which implies that the distributions of bids for both genders are not significantly different from each other.



Figure 2: Histogram of bids to earn money for the group by gender