Sep, 2025
GRINS DISCUSSION PAPER SERIES DP N° 04/2025 ISSN 3035-5576

GRINS

Perceptions of Gender Norms: Framing Effects and Double Standard

DP N° 04/2025

Authors:
Francesca Barigozzi, Caterina Gaggini, Natalia Montinari

Sep, 2025
GRINS DISCUSSION PAPER SERIES DP N° 04/2025 **ISSN 3035-5576**

Perceptions of Gender Norms: Framing Effects and Double Standard

Francesca Barigozzi, Caterina Gaggini, Natalia Montinari

KEYWORDS

Gender norms | Krupka&Weber elicitation method | Representative surveys | Domestic Chores | Framing effects | Double Standard

JEL CODE

A13, C90, D01, J16

ACKNOWLEDGEMENTS

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Part of this study was funded by the European Union - NextGenerationEU, in the framework of the "GRINS -Growing Resilient, INclusive and Sustainable project" (PNRR - M4C2 - I1.3 - PE00000018 – CUP J33C22002910001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. We thank the participants of the 2022 ESA Conference (Bologna, Italy), the 2024 ESPE Conference (Rotterdam, Netherlands), 2024 NSILE (Aarhus, Denmark). The paper was also presented at the seminars of the University of Granada (Spain, June 2024), and Aarhus University (Denmark, September 2024), we thank all participants for their useful comments.

CITE THIS WORK

Author(s): Francesca Barigozzi, Caterina Gaggini, Natalia Montinari. Title: Perceptions of Gender Norms: Framing Effects and Double Standard. Publication Date: 2025.

To what extent is domestic labor still perceived as a female responsibility? Do traditional gender norms contribute to its persistently unequal division within households? To answer these questions, we design a survey experiment where an incentivized measure of normative expectations (Krupka and Weber, 2013) is embedded within a representative survey of the Italian population (N=1,501). In this way, we bridge two strands of literature: survey-based elicitation of attitudes and incentivized experimental measures of social appropriateness. Participants evaluate the social appropriateness of chore allocations in vignettes where partners' labor supply, household division, and the gender of the proposer vary. We show that, when partners have the same

GRINS DISCUSSION PAPER SERIES DP N° 04/2025 ISSN 3035-5576

working status, equal sharing of household chores is widely recognized as socially appropriate across generations. However, judgments of unequal allocations reveal the presence of a framing effect and a gender double standard, especially among middle and older generations. Younger generations exhibit greater internalization of egalitarian norms, suggesting an ongoing shift in attitudes. Finally, we find that perceived norms on the division of household labor, measured through normative expectations, are strongly associated with women's labor market outcomes at the regional level. These findings highlight the cognitive biases sustaining gender inequality inside and outside the household.|

Keywords: Gender norms, Krupka&Weber elicitation method, Survey Experiments, Domestic Chores, Framing effects, Double Standard.
JEL classification: A13; C90; D01; J16.

Perceptions of Gender Norms: Framing Effects and Double Standard*

Francesca Barigozzi†, Caterina Gaggini‡
and Natalia Montinari§

September 2025

Abstract

To what extent is domestic labor still perceived as a female responsibility? Do traditional gender norms contribute to its persistently unequal division within households? To answer these questions, we design a survey experiment where an incentivized measure of normative expectations (Krupka and Weber, 2013) is embedded within a representative survey of the Italian population (N=1,501). In this way, we bridge two strands of literature: survey-based elicitation of attitudes and incentivized experimental measures of social appropriateness. Participants evaluate the social appropriateness of chore allocations in vignettes where partners' labor supply, household division, and the gender of the proposer vary.

We show that, when partners have the same working status, equal sharing of household chores is widely recognized as socially appropriate across generations. However, judgments of unequal allocations reveal the presence of a framing effect and a gender double standard, especially among middle and older generations. Younger generations exhibit greater internalization of egalitarian norms, suggesting an ongoing shift in attitudes. Finally, we find that perceived norms on the division of household labor, measured through normative expectations, are strongly associated with women's labor market outcomes at the regional level. These findings highlight the cognitive biases sustaining gender inequality inside and outside the household.

Keywords: Gender norms, Krupka&Weber elicitation method, Survey Experiments, Domestic Chores, Framing effects, Double Standard.
JEL classification: A13; C90; D01; J16.

*Part of this study was funded by the European Union - NextGenerationEU, in the framework of the "GRINS -Growing Resilient, INclusive and Sustainable project" (PNRR - M4C2 - I1.3 - PE00000018 – CUP J33C22002910001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

We thank the participants of the 2022 ESA Conference (Bologna, Italy), the 2024 ESPE Conference (Rotterdam, Netherlands), 2024 NSILE (Aarhus, Denmark). The paper was also presented at the seminars of the University of Granada (Spain, June 2024), and Aarhus University (Denmark, September 2024), we thank all participants for their useful comments.

†Correspondent Author: University of Bologna, Department of Economics, Piazza Scaravilli 2, Bologna, Italy, Email: francesca.barigozzi@unibo.it.

‡Aarhus University, Department of Economics and Business Economics, Fuglesangs Allé 4, 2620, 201, Aarhus V, Denmark, Email: caterinagaggini@econ.au.dk.

§University of Bologna, Department of Economics, Piazza Scaravilli 2, Bologna, Italy, Email: natalia.montinari2@unibo.it.

1 Introduction

Gender norms change slowly, and despite the recent improvements in women's labor market prospects, society still maintains different expectations for women and men. Childcare and household chores remain predominantly female tasks, whereas men are expected to invest primarily in their careers. Italy stands out as a negative benchmark in official statistics (OECD, 2019) and in comparative studies on gender gaps in time use based on time-diary surveys. Specifically, Italy presents one of the largest gender gap in time devoted to informal childcare and household work along all stages of the life course; see, among others, Anxo et al. (2011); Craig and Mullan (2011); Gimenez-Nadal and Molina (2020). According to the Harmonized European Time Use Survey (HETUS), in 2010, the Italian gender gap in time spent on household and family care activities was 2.47 daily hours. This gap was not balanced by the gap in paid work, which Eurostat estimated to be 1.52 daily hours in 2010. Using ISTAT (Italian Institute of Statistics) *Use of Time* surveys, Cappadozzi (2019) and Barigozzi et al. (2023) report that in 2014, the gender gap in total work (paid and unpaid) among parents in dual-earner couples, where both partners worked full time, exceeded one hour per working day (equivalent to 11 hours per week). As a potential explanation for these gaps, the authors of all mentioned studies highlight Italy's strong social norms regarding gender roles.

This pattern is confirmed by more recent data, such as those collected through our nationally representative survey conducted in 2020 (see below for details). Table 1 presents results for the subsample of respondents in a couple who report how household tasks are divided between partners in their household. The modal response indicates that routine domestic tasks—such as cooking, cleaning, and laundry—are predominantly performed by women, with shopping representing a notable exception. This suggests that domestic labor continues to be predominantly perceived as a female responsibility, even in dual-earner households and across generations (see Appendix for additional results).[1]

Is the observed unequal division of domestic duties attributable to the persistence of traditional norms on gender roles?

[1] In the Appendix, we replicate the analysis restricting the sample to respondents who are both partnered and employed (Table A1), and explore heterogeneity by age group (Table A2). The results confirm the robustness of the patterns documented here.

"Who usually performs the following household tasks in your couple?"

Panel a): Female Respondents living in couple, N=700

	Wash and Iron	Shopping	Cooking	Cleaning
Always me or mostly me	84.38	61.71	75.16	76.88
Equally	12.02	29.76	16.67	19.57
Always or mostly my partner	1.61	8.53	8.01	1.68
Another person	1.99	0.00	0.16	1.87

Panel b): Male Respondents living in couple, N=493

	Wash and Iron	Shopping	Cooking	Cleaning
Always me or mostly me	34.81	42.73	34.60	34.87
Equally	18.75	36.84	24.46	25.91
Always or mostly my partner	43.80	20.23	40.30	36.50
Another person	2.63	0.19	0.64	2.71

Table 1: **Subsample of married or cohabiting individuals by gender.** Distribution of responses regarding how household tasks are divided between respondents and their partners. The sample (N=1,501) was collected in June 2020 by the professional company Scenari S.r.l. in June 2020 from a panel of 10, 000 participants using the computer-assisted web interviewing (CAWI) methodology. The sample is representative with respect to gender (male, female), age range (25-34; 35-49; 50-64), and residence area (North, Center, and South of Italy). The table displays the eighteen groups relevant to our social norm elicitation Percentages are computed using Stata's `svy` prefix, which accounts for sampling weights and complex survey design. Statistical significance of gender differences in response distributions was assessed using the Rao–Scott adjusted chi-square test (design-based F-test). For all tasks, the gender differences in response distributions are statistically significant at the 1% level ($p < 0.001$). All reported p-values were adjusted for multiple comparisons using the Benjamini-Hochberg False Discovery Rate correction (Benjamini & Hochberg, 1995).

To investigate social norms related to the division of household duties between partners, we designed a survey experiment and administered it to a representative sample of the Italian population (N = 1,501). Drawing on the Krupka & Weber approach (Krupka and Weber (2013), KW henceforth), we design two vignettes specifying the partners' working status and three possible allocations of household work—equal and unequal. We find that an egalitarian norm regarding the division of household duties prevails when respondents evaluate the vignette where partners share the same employment status. In other words, equal sharing of domestic work between partners working full-time is perceived as socially appropriate by the large majority of respondents.

How can the existence of an egalitarian social norm be reconciled with the persistent inequality in real-world behaviors documented by Time Use diary studies and our own survey? Using the KW methodology, we explore this puzzle by asking respondents to evaluate the scenarios in which the division of household duties deviates from an equal split, systematically varying the gender of the partner initiating the allocation.

This design, applied to the vignette featuring partners with the same working conditions, reveals two distinct biases: a *framing effect* and a *double standard*. The *framing effect* arises when the same unequal proposal is evaluated differently depending on whether it is initiated by the man or the woman. A *double standard* emerges when a division favoring the woman is judged more negatively than an equivalent one favoring

the man. These unconscious biases—*framing effects* and *double standards*—may help explain the tension between stated egalitarian norms and actual behaviors, especially for the older generations of our sample. These biases likely influence how individuals perceive and justify their own daily contributions to domestic work.

We also analyze a second vignette featuring unequal working conditions—specifically, a scenario in which the male partner works full-time, the female partner part-time, and the male partner is described as the primary contributor to household labor income. Applying the KW methodology to this setting, we investigate the perceived social appropriateness of an equal division of household tasks across different age groups, in order to assess generational differences in gender norm perceptions.

From a methodological perspective, our approach innovates relative to the existing literature by embedding the KW incentivized elicitation of normative expectations within a representative survey of the Italian population. In doing so, we bridge two strands of research: (i) incentivized experimental measures of social appropriateness and (ii) survey-based elicitation of values and attitudes. This integration addresses key limitations of both approaches—namely, the lack of representative subject pools in experimental studies, and the limited ability of standard surveys to capture shared prescriptive norms through sophisticated elicitation techniques. Section 1.1 further elaborates on our methodological contribution.

According to the KW methodology, we elicit social norms on gender roles as incentivized modal responses in a coordination game. In the latter, we ask respondents to match the choice of a group of people similar to them regarding gender, age, and residence area. In such a way, we create homogeneous subgroups in which respondents guess modal responses. As a coordination device, respondents should use the *shared* social norms, or their perception of what is socially appropriate in their group. In addition, in our analysis, we control for the personal characteristics of the respondents (e.g., civil status, education, employment status, presence of children, etc.) and personality traits.

Our sample is representative in terms of the three key characteristics that define the groups analyzed: gender, age, and area of residence. These characteristics are likely to significantly shape perceptions of gender norms (see Section 2.1 for discussion). Regarding the age of the respondents, the sample is representative of three age groups: 25–34, 35–49, and 50–64. These cohorts allow us to assess whether, and in what ways, elder groups hold more conservative gender norms compared to younger groups. In other words, comparing social norms elicited from groups of different ages offers valuable insight into the evolution of gender norms in society.

Our empirical hypotheses are formalized following a theoretical model in which partners contribute time to a family public good and face disutility when deviating from a shared norm over the socially approved division of domestic chores (see Section 4).

We use vignettes to elicit respondents' opinions for several reasons. First, vignettes provide a standardized scenario that all respondents consider, ensuring comparability of responses across individuals. Second, they offer a context that helps the respondents

understand abstract concepts. Third, by exogenously varying the working arrangements of the partners in our vignettes, we can analyze the causal impact of this variation on respondents' judgments. Finally, vignettes can reduce social desirability bias because respondents are asked to comment on a hypothetical situation and not to report on their personal choices.

Describing our results in more detail, we find that when partners have the same working status, the equal sharing of household chores is broadly acknowledged across generations as the socially appropriate behavior. However, two biases—(i) a framing effect and (ii) a double standard—influence the assessment of deviations from the equality norm. While the two older generations are affected by both biases, younger respondents are not subject to the framing effect and exhibit a significantly weaker double standard. Additionally, we find (iii) evidence of a decline in the "male as the breadwinner" model among young adults. Hence, our findings suggest that young adults are genuinely internalizing more egalitarian norms.

Finally, we show that (iv) perceived social norms display a significant association with women's labor market outcomes based on administrative data at the regional level. This suggests that perceived norms on the division of domestic labor have external validity in explaining labor market outcomes and, they help account for inequality not only within the household but also in the labor market.

Results (i) and (ii) are obtained by focusing on the vignette where partners have the same labor market status. Result (i) indicates that the gender of the partner proposing a chores allocation significantly influences perceptions of social appropriateness. Specifically, in middle and older generations, a woman who proposes a self-beneficial distribution of chores is stigmatized more than a man making the same proposal.

Regarding result (ii), we find that when the woman proposes the allocation, deviations from the equality norm are judged asymmetrically, especially by the middle and older generations, providing evidence of a double standard. In other words, a woman who offers to contribute less than her partner to household chores is rated as less appropriate than when offering to contribute more, whereas the same pattern does not hold when the man proposes the chores allocation.

We thus document two biases in how chore allocations are evaluated, both reinforcing implicit gendered expectations. These biases arise despite a prevailing social norm that favors an equal division of tasks between partners with similar working arrangements, helping to explain why time-use data continue to show persistent gender differences in the division of family chores.

Result (iii) is obtained from the vignette in which partners have unequal labor market participation, with the female partner working part-time. In this setting, we find that older generations are less likely than younger ones to perceive an equal division of household tasks as socially appropriate.

Finally, regarding result (iv), and inspired by Fortin (2005), we examine the explanatory power of the normative expectations elicited in our representative survey. Specif-

ically, we study the association between these elicited social norms and female labor market participation across Italian regions. We find a positive relationship between our measure of social norms and women's labor market participation at the age-group and regional level, providing further evidence that perceived gender norms play a meaningful role in shaping economic behavior.

The rest of the paper is organized as follows: Section 1.1 explains our methodological contribution, Section 1.2 discusses our contribution to the growing empirical literature on gender norms. Section 2 describes the survey and the experimental treatment; Section 3 presents our hypotheses, Section 4 sketches a theoretical model. Finally, Section 5 presents the results, and Section 6 concludes.

1.1 Our methodological contribution

As mentioned before, we bridge two strands of literature on social norms: (i) incentivized experimental measures of social appropriateness and (ii) survey-based elicitation of values and attitudes.

Ostrom (2000) defines social norms as "shared understandings about actions that are obligatory, permitted, or forbidden" (page 143–144). This definition highlights a key feature of social norms. They must be *jointly recognized by a group*. In line with this definition, recent experimental literature has employed coordination games, conducted either in the field or the laboratory, to elicit social norms shared at the group level (see KW and references therein). In these experiments, participants are given monetary incentives to match the responses of others. As a result, they play a matching coordination game in which a natural strategy is to align with expectations about what most people consider socially appropriate or inappropriate in the described context. In this way, coordination games serve to elicit beliefs that are held at the group level. Specifically, in KW, people's beliefs about others' beliefs are *higher-order beliefs* elicited at the group level.[2] This elicitation method has the advantage of aligning with Ostrom's idea that the *collective* approval or disapproval of certain behaviors within a specific group is at the very core of the definition of social norms.

Shifting to the the survey-based elicitation of attitudes, the misperception of gender norms has become a central topic in recent empirical literature on social norms; see Bursztyn et al. (2020), Bursztyn et al. (2023), Bursztyn and Yang (2022), Boneva et al. (2024), and Cortés et al. (2024). While these studies focus on (incorrect) expectations about others' beliefs rather than explicitly defining social norms, they tend to identify the actual social norm as the dominant personal value within a given population, operationalized as the average first-order belief.[3] A limitation of this (albeit implicit)

[2]Conversely, in the recent empirical literature on social norms, beliefs about others' beliefs are *second-order beliefs* because first-order beliefs are elicited first, and respondents are then asked to estimate these previously elicited first-order beliefs; see the survey by Nosenzo and Görges (2020) (page 288).

[3]Cortés et al. (2024) write on page 1: "The scenarios ask respondents about (1) their own recommendation (first-order beliefs) and their perceptions of the recommendations of those living in the same part of the country as them (second-order beliefs) about whether a mother with a young child should accept a job offer to return to work." And then, on page 11, "To study misinformation, we compare

operationalization of social norms in the survey-based empirical literature is that it does not incorporate the idea that social norms should reflect expectations *shared by the reference group*. Moreover, the fact that individuals' first-order beliefs are influenced by information about others' beliefs—as shown by Bursztyn et al. (2020), Cortés et al. (2024), Bursztyn et al. (2023), and Boneva et al. (2024)—suggests that associating the actual social norm with prevalent first-order beliefs might be problematic. These studies show that when individuals are informed about others' expectations, they tend to adjust their own first-order beliefs, indicating that the social dimension of norms is crucial for their measurement. This underscores that social norms should encapsulate the shared perception of a group's normative values.

However, the experimental literature also presents limitations. Specifically, the group of participants in lab experiments is necessarily small and, being predominantly composed of university students, may not be representative of society as a whole, particularly for topics such as gender norms. This shortcoming has been recently mitigated by online studies and field experiments.

To address the limitations of both experimental studies and survey-based elicitation of attitudes, we adopt the KW experimental methodology to elicit social norms but apply it to a representative sample of the Italian population. By embedding this experimental methodology into a large-scale, representative survey, we combine the theoretical rigor of KW's approach—preserving the definition of social norms as collective expectations—with the ability to measure social norms at the population level. This integration offers a cost-effective, behaviorally validated method that retains the strengths of experimental economics while leveraging a nationally representative sample to capture social norms across an entire country.

1.2 Related literature

Our study builds on the experimental literature that employs KW's methodology to elicit social norms by incentivizing beliefs held at the group level. Social norms measured through this methodology have been shown to predict behavior in various settings, including prosocial behavior, bribery, discrimination, and saving behavior; see among others, Gächter et al. (2013); Burks and Krupka (2012); Barr et al. (2018); Fromell et al. (2021). While prior research has predominantly applied this approach in controlled laboratory experiments or specific field environments, our study expands the scope by applying KW's methodology to a nationally representative sample of the Italian population, thereby providing broader insights into how social norms operate in a generalizable real-world context. This extension to a representative sample allows us to address questions that have remained unexplored in previous studies focused on more homogeneous or experimental populations. Specifically, by capturing social norms within a diverse and demographically varied population, we gain insights into the heterogeneity of perceived

an individual's second-order beliefs with the average first-order beliefs of people of their gender in their state."

norms across different socio-economic and cultural groups. Moreover, our analysis contributes to the understanding of how social norms evolve and differ at the population level, as opposed to the more localized and specific settings typically analyzed in the experimental economics literature.[4]

Our paper is also related to the literature examining the relationship between gender norms and women's economic outcomes, aimed at understanding whether social norms constrain women's labor market choices. Fortin (2005) uses the World Values Survey (WVS) to analyze the impact of attitudes toward gender roles, competition, and various aspects of work on women's employment decisions and part-time status among employed women.[5] Similarly, Fernández and Fogli (2005), Bertrand et al. (2015), Fortin (2005), Kleven et al. (2019), and Bertrand et al. (2021) examine the association between labor market outcomes and agreement with statements from representative surveys such as the WVS, the European Values Survey, the International Social Survey Programme (ISSP), or the International Values Survey. Section 5.3 compares the explanatory power of social norms regarding gender roles elicited through KW's methodology with social norms measured via agreement with statements from representative surveys.

Additionally, we share a specific focus on the evolution of gender norms in society with Fortin (2005) and Bertrand et al. (2021). However, unlike those papers, which address the issue by comparing subsequent waves of the same survey, we analyze three different age groups interviewed in our survey. As mentioned in the Introduction, our survey is representative also concerning three age ranges of respondents (25–34, 35–59, and 50–64), enabling us to disaggregate and compare their responses based on age.

Our study is also related to the recent literature on the misperception of gender norms, particularly Bursztyn et al. (2020) and Cortés et al. (2024). Bursztyn et al. (2020) investigate the prevailing gender norm among Saudi Arabian men regarding women working outside the home. They ask a sample of Saudi Arabian men whether they agree or disagree with the statement: "*In my opinion, women should be allowed to work outside of the home.*" Participants are then asked, and incentivized, to estimate the percentage of other participants who agree with the statement, providing a measure of misperception of the social norm. Although both our study and Bursztyn et al. (2020) involve eliciting beliefs about others' beliefs, our approaches to the definition of social norms differ. Bursztyn et al. (2020) implicitly defines social norms as the prevalent first-order belief (i.e., the dominant personal value, either agree or disagree). In contrast, we adopt KW's approach, defining a social norm as the mode of higher-order beliefs. Regarding objectives, Bursztyn et al. (2020) focus on the misperception of the

[4]While we acknowledge that beliefs about others' beliefs may be influenced by gender stereotypes, we do not address gender stereotypes explicitly. See Bordalo et al. (2019) for laboratory experiments that explore how gender stereotypes shape beliefs about the ability of oneself and others in different categories of knowledge.

[5]Specifically, agreement with the statement "When jobs are scarce, men have more right to a job than women" stands out as the most powerful explanatory factor of cross-country differences in female employment rates and the gender pay gap. This statement captures the perception of the man as the breadwinner, as well as discriminatory attitudes against working women. Agreement with the statement "A working mother can establish just as warm and secure relationship with her children as a mother who does not work" is closely associated with women's employment status and mother's guilt.

gender norm regarding women working outside the home among young men in Saudi Arabia and study how information can serve as a policy intervention against conservative norms. Our study, in contrast, aims to analyze the evolution of gender norms in a representative survey of the Italian population using vignettes that offer standardized scenarios, ensuring contextualization and comparability across individuals.

Similarly to Bursztyn et al. (2020), Boneva et al. (2024) provide cross-country evidence on the persistence of gender norms driven by systematic misperceptions. Using a representative survey across six countries, including Italy, they show that while most men prefer an equitable division of household tasks, this preference is consistently underestimated by both men and women. Their study demonstrates that correcting these misperceptions through an informational intervention significantly shifts beliefs and increases self-reported support for gender equity within couples. These findings align with our analysis of generational shifts in normative expectations and suggest that pluralistic ignorance may be a key factor in the persistence of traditional gender roles. Our study focuses specifically on generational shifts in gender norms within Italy. While Boneva et al. (2024) test an informational intervention, we use incentivized coordination games to measure normative expectations and examine their link to labor market outcomes.

The study most closely related to ours is Cortés et al. (2024). They explore how second-order beliefs shape first-order beliefs using two vignettes and an informational treatment presented to a representative sample from the New York Fed's Survey of Consumer Expectations. For the first vignette, respondents are asked about their second-order beliefs regarding the perceived appropriateness of "A mother with a preschool child working when her husband has a job, she receives a job offer she likes and pays well, and a high-quality, free public pre-kindergarten is available." Half of the respondents are then given information about second-order beliefs of other respondents of the same gender and state of origin before being asked about their own first-order beliefs. The second vignette is similar but considers high and low opportunity costs of the mother receiving the job offer. Cortés et al. (2024) primarily aims to understand the role of misperceptions and information gaps in the persistence of gender norms in the U.S. In contrast, our study compares gender norms across three representative subsamples with different ages to trace the evolution of norms. Similar to Bursztyn et al. (2020), Cortés et al. (2024) implicitly defines (actual) social norms with the prevalent first-order belief, while we use the mode of higher-order beliefs. Nevertheless, our study shares methodological similarities with theirs, as both papers present two vignettes to a representative sample.

Finally, Barigozzi and Montinari (2025) analyze data from the same representative survey used in this paper. They compare two methodologies for measuring social norms: KW's experimental approach (the mode of incentivized higher-order beliefs) and the approach commonly used in the empirical literature (the prevalent first-order belief). They examine two prescriptive statements, i.e., "When jobs are scarce, men should have more rights to a job than women," and "A woman should be ready to reduce the time devoted to her job for family reasons." Barigozzi and Montinari (2025) show

that analyses based on personal values produce a significantly more progressive proxy of gender norms than those elicited through coordination games. Specifically, they find that most respondents report first-order beliefs that are more progressive than higher-order beliefs. This behavior occurs regardless of whether respondents correctly perceive others' beliefs and is positively associated with holding a university degree. It may reflect genuinely progressive views, a bias toward appearing progressive, or an ongoing shift in social norms. Overall, this paper suggests that the risk of noisy elicitation of social norms due to social desirability bias remains high in those studies that identify social norms with first-order beliefs; and more so when social norms are changing relatively fast like gender norms.

2 The representative Survey

We designed a survey that provides incentivized elicitation of social norms over possible action choices determining different degrees of gender equality in the allocation of housework between two partners of opposite sex. We collected data on a representative sample of the Italian population (N=1,501).[6] Representativeness holds with respect to the following characteristics: gender (male, 41.57%; female, 58.43%), age range ($25-34$ (19.85%); $35-59$ (52.43%); $50-64$ (27.71%)), residence area (North (47.90%), Center (18.92%) and South of Italy (33.18%)) and, education (percentage of people holding a tertiary degree: 35.38%), see Table 2. Descriptive statistics are provided in Tables A3 in the Appendix, while a comparison of our dataset with data from ISTAT (2019) is provided in Table OA1 of the Online Appendix.

The data was collected by the professional company Scenari S.r.l. in June 2020 from a panel of $10,000$ participants using the computer-assisted web interviewing (CAWI) methodology.[7] On average, participants spent 23.4 minutes completing the survey (standard deviation: 29.83 min).

Note that we used a commercial survey company that employs quota-sampled panels, a common approach in survey research; see, among others, Stantcheva (2023). While this method allows for a good approximation of population characteristics based on observable variables, we acknowledge that there may be self-selection in the decision to enroll in the panel. However, as with all non-probability sampling methods, there may be dimensions in which our sample is not fully representative, a common issue for research utilizing survey experiments; see, among others, Alesina et al. (2023) and Settele (2022).

The survey is organized in 3 parts (see Table 3): in the first part, participants answered questions on their demographic information and household composition. In the second part, we elicited social norms following the methodology introduced by KW; we proposed four vignettes and a question composed of five claims to measure social norms and

[6] The size of our sample is in between the two most recent waves of the WVS for Italy, i.e. wave 5 ($N = 1,012$) and wave 7 ($N = 2,282$).

[7] CAWI is an internet surveying technique whose main advantage is to have a lower cost compared to other methods, basically because there is no need for interviewers to hold the survey.

	North		Center		South and Islands	
Age group	Male	Female	Male	Female	Male	Female
Age 25-34	63	67	20	26	58	64
Age 35-49	133	244	68	92	105	145
Age 50-64	91	121	32	46	54	72
Total	287	432	120	164	217	281
N (M+F)	719		284		498	

Table 2: **Groups size in the representative sample** (N=1,501).

Note: The sample (N=1,501) was collected in June 2020, it is representative with respect to gender (male, female), age range (25-34; 35-49; 50-64), and residence area (North, Center, and South of Italy). The table displays the eighteen groups relevant to our social norm elicitation.
The North includes the regions of the North-West (Liguria, Lombardy, Piedmont, Aosta Valley) and those of the North-East (Emilia-Romagna, Friuli Venezia Giulia, Trentino-Alto Adige, Veneto). The Center includes the regions of Lazio, Marche, Tuscany, and Umbria. The Mezzogiorno (South and Islands) includes the regions of Southern Italy (Abruzzo, Basilicata, Calabria, Campania, Molise, Apulia) and the insular regions (Sardinia, Sicily).

personal values.[8]

For each of the four vignettes and each of the five claims, participants were asked to guess the answer chosen by the majority of other respondents similar to them with respect to gender, age group, and residence area, i.e., their higher-order beliefs. The four vignettes were presented randomly but always before the question containing the claims. Participants were unable to go back to previously answered questions, and they were unaware of the content of the different parts of the survey.

The four vignettes differ along two dimensions (within-subject variation): i) the presence of children or not, ii) whether the two partners have or not the same working status. In addition, we varied, between subjects, the gender of the partner who proposes the allocation of the chores. More details on the vignettes and the social norms elicitation are provided in the next section.

The company offers incentives to motivate members of the panel to take part in surveys adopting a point-based system. Participants receive points for each survey they complete, depending on the survey length. Every 50 points they can get a 10 Euros Amazon gift card. For our survey, the company offered 20 points; in part 2, we provided additional incentives as part of the (incentivized) norm elicitation: participants who correctly guessed the answer given by most individuals in their reference group were rewarded with 3 Euros per correct guess paid for with an Amazon gift card. At the beginning of part 2 participants were informed that after the completion of the data collection, one of the questions presented in part 2 as well as 10% of participants (i.e. $N = 150$) would be randomly selected to receive the earnings associated with their correct guesses.[9,10]

[8]The two vignettes involving a child are not analyzed in this paper, so we avoid going into detail about them. The results are partially replicated and available on request. The five claims are not included in this study. Some of them are analyzed in Barigozzi and Montinari (2025).

[9]A translation of the explanations shown to the participants is presented in the Online Appendix Table OA2.

[10]Charness et al. (2016) provide evidence that paying for only a subset of individuals or for a subset

Survey sections	
Part 1	Demographic and household composition
Part 2	Incentivized norms elicitation following Krupka and Weber (2013)
Part 3	Chores allocation in the household
	Personal values (unincentivized)
	Employment, political orientation, personality traits,...

Table 3: **Survey sections**.

In the third part, participants answered questions about i) their employment, and the employment of other members of their household; ii) the allocation of the chores within their household (before, during, and after the lockdown associated with the first wave of the COVID-19 emergency); iii) their (unincentivized) personal values on the same questions encountered in part 2 (i.e. the vignettes, and the question with the five claims); iv) their political orientation, the relative importance of different spheres of life (e.g. family, work, friends); v) some personality traits (TIPI, Gosling et al. (2003), cognitive reflection tests, Frederick (2005)).

2.1 Gender Norms Elicitation

Participants were presented with a set of vignettes depicting a hypothetical situation where one of the partners of a fictional couple chooses how to divide household chores; see the bottom part of the Online Appendix Table OA2. We focus on the *Full-Time* and *Part-Time* vignettes, reproduced in Tables 4 and 5, respectively. In both cases, the male partner works full-time, while the female partner's labor market participation differs: she works full-time in the former and part-time in the latter.[11]

Respondents were randomly exposed either to the version of the Vignettes where the female partner is proposing the chores allocation (54.26%) or to the version where the male partner is proposing the allocation (45.74%). In other words, we vary between subjects the gender of the partner proposing the housework sharing. As we explain when stating our hypothesis, we expect that the identity of the partner proposing the allocation significantly influences how the allocation is perceived.

Table A4 in the Appendix controls that randomization worked by testing differences by proposer's gender in our variables of interest.

Table 5 and Table 4 present the woman (man) proposing versions.

Respondents are asked to judge three scenarios within each vignette. In the first scenario, the female (or male) partner is willing to do most of the household work. In the second, partners share the household work equally. In the third, the female (or male) partner is

of decisions is as effective as the "pay all" approach. See also Burks and Krupka (2012) who ran a social norm elicitation and randomly selected 25% of participants for the payment of the social norm elicitation task. Eventually, one of the four vignettes was randomly selected for payment. Of the 150 participants randomly selected, 39% provided 2 correct answers out of 3 in the vignette, earning on average 5 Euros, for a total cost of 745 Euros, paid for incentives.

[11] We could have included more detailed descriptions in the vignettes (e.g., specifying which household chores are involved in task-sharing or whether partners share their income). However, we deliberately chose not to ask for judgments on such highly specific scenarios to avoid making the vignette overly complex and narrowly focused.

Vignette Full-Time: Equality between partners	
Antonio and Francesca are either married or cohabiting partners. They both work the same number of hours, earn roughly the same amount of money, and have similar career trajectories. They have no children and no one to help them with household chores.	
Questions	
How would most people similar to you (i.e., of your same gender, age group, and living in the same geographic area) evaluate Francesca (Antonio)'s behavior in the following scenarios?	
$V1_1$	Francesca (Antonio) is willing to take care of up to $\frac{1}{4}$ ($\frac{3}{4}$) of the household chores and leaves $\frac{3}{4}$ ($\frac{1}{4}$) of them to Antonio (Francesca).
$V1_2$	Francesca (Antonio) is willing to evenly split the household chores with Antonio (Francesca).
$V1_3$	Francesca (Antonio) is willing to take care of up to $\frac{3}{4}$ ($\frac{1}{4}$) of the household chores and leaves $\frac{1}{4}$ ($\frac{3}{4}$) of them to Antonio (Francesca).
Possible answers	Definitely Inappropriate, Somewhat Inappropriate, Somewhat Appropriate, Definitely Appropriate

Table 4: **Text of Vignette Full-Time depicting equality between partners**.

willing to take on only a small share of the household work.

To elicit gender norms, respondents were asked to rate the social appropriateness of every household chore allocation as they thought their reference group would. According to the KW methodology, this creates a coordination games where social norms represents a salient coordination device: in order to match the ratings of others, individuals are likely to rely on their beliefs about what most people consider appropriate or inappropriate. If all participants follow the same reasoning, the KW methodology reveals the social norm shared in the group.
Specifically, respondents were asked to guess how most people in their reference group would evaluate the social appropriateness of each allocation using a four-point Likert scale (Very Inappropriate, Somewhat Inappropriate, Somewhat Appropriate, Very Appropriate). Following KW, we did not include a neutral option on the Likert scale as this would result in the risk of respondents using the neutral point as a coordination device (instead of the norm).

A reference group is a set of people characterized by the same gender (male, female), age range (25-34; 35-59; 50-64), and residence area (North, Center, and South and Islands of Italy), and respondents are recalled their reference group before the elicitation. The fact that groups are contingent on gender is quite natural, given our focus on gender norms. For example, respondents may think that men hold more conservative beliefs than women on the role of women in society. In addition, groups are contingent on the respondents' age because younger people might hold less conservative beliefs than older people. Likewise, it has been observed that new generations tend to be more progressive than older ones, as respondents' replies in older and more recent waves of the WVS indicate (see, among many others, Fortin 2005). Finally, our groups are contingent on the region where respondents live because it has been shown that social norms differ substantially in Italy between the North and South, with residents in the South being more conservative than those in the North (see, e.g., Del Boca 2002 and Bigoni et al. 2016).

To sum up, participants play a pure matching coordination game whose goal is to

Vignette Part-Time: Asymmetry between partners	
Imagine Giulio and Silvia: they are either married or cohabiting. Giulio works twice as many hours as Silvia and earns about twice as much. They have no children and no one to help them with household chores.	
Questions	
How would most people similar to you (i.e., of your same gender, age group, and living in the same geographic area) evaluate Silvia (Giulio)'s behavior in the following scenarios?	
$V1_1$	Silvia (Giulio) is willing to take care of up to $\frac{1}{4}$ ($\frac{3}{4}$) of the household chores and leaves $\frac{3}{4}$ ($\frac{1}{4}$) of them to Giulio (Silvia).
$V1_2$	Silvia (Giulio) is willing to evenly split the household chores with Giulio (Silvia).
$V1_3$	Silvia (Giulio) is willing to take care of up to $\frac{3}{4}$ ($\frac{1}{4}$) of the household chores and leaves $\frac{1}{4}$ ($\frac{3}{4}$) of them to Giulio (Silvia).
Possible answers	Definitely Inappropriate, Somewhat Inappropriate, Somewhat Appropriate, Definitely Appropriate

Table 5: **Text of Vignette Part-Time depicting asymmetry between partners**.

anticipate the extent to which other participants similar to them will rate scenarios as socially appropriate or inappropriate. This implies that we elicit respondents' higher-order beliefs. Then, following KW, we define social norms as the mode of the distribution of higher-order beliefs reported by members of a group on a specific scenario for each vignette.

Note that each participant encountered each vignette twice, first in part 2 (where incentivized higher-order beliefs are elicited) and then in part 3 of the survey (where unincentivized first-order beliefs or personal values are elicited). We only implement one sequence of elicitation, collecting the incentivized measures first and then the unincentivized ones.[12]

3 Hypotheses

In this section, we present our main hypotheses.

Our first hypothesis is that proposing both an advantageous and a disadvantageous chore allocation is evaluated differently based on the proposer's gender. In other words, we expect that framing effects shape perceptions of social appropriateness:

Hypothesis 1. ***Framing effects.*** *The gender of the individual initiating the allocation of chores influences the perceived social appropriateness of the proposal. Specifically, a woman suggesting an allocation that benefits herself at her partner's expense is judged more harshly than a man in an equivalent scenario.*

To test this hypothesis, we focus on Vignette Full-Time, where both partners have similar working arrangements. In the simple model presented in Section 4, we consider a unitary couple whose partners maximize their joint utility by contributing to the

[12]Robustness of KW's method with respect to the order of elicitation of first and higher-order beliefs is reported by König-Kersting (2021), along with more general evidence of the robustness of this methodology to several variations: i.e. to the timing of play of the game with respect to the elicitation (d'Adda et al. 2016) and to the interests at the stake of the respondent (i.e. stakeholder or spectator, Erkut et al. 2015).

household's public good, simultaneously deciding how much time to devote to household chores. Featuring Vignette Full-Time, given the partners' similar working conditions, we assume that they experience the same disutility when allocating time to household chores; see Section 4.1. Additionally, it is reasonable to assume that the social norm dictates an equal contribution from both partners to the family's public good. This assumption is indeed supported by Table 6 below, which shows that, in the social norm elicitation task, the modal response for the "equal contribution" scenario is *very appropriate*.

In our model, deviating from this egalitarian norm generates a disutility, inspired by Fehr and Schmidt (2006), where contributing less than the egalitarian norm is perceived as more inappropriate than contributing more. Notably, we expect the gender of the proposer to shape perceptions of appropriateness due to implicit biases on household responsibilities. The literature on framing effects suggests that identical information may be judged differently depending on its source (Tversky and Kahneman 1981; Carpenter 2022). In our setting, this implies that identical deviations from equal chore sharing may be evaluated differently depending on whether the proposal comes from the male or the female partner; see Definition 1 at the end of Section 4.1.

In other words, if individuals hold gendered priors, they may evaluate a scenario in which a woman proposes a self-beneficial allocation as more inappropriate than the mirror-image scenario where her partner proposes the self-beneficial allocation. We will also assess whether framing effects exist across age groups.

Our second hypothesis speculates on the existence of a gender double standard.

Hypothesis 2. ***Gender Double Standard:*** *Women are judged as more socially inappropriate than men for self-beneficial deviations from the equal contribution to household duties.*

To test this hypothesis, we focus again on Vignette Full-Time and analyze the perceived social appropriateness of deviations from the equal-contribution scenario, depending on the gender of the partner benefiting from the deviation, as illustrated in our theoretical model below. If a gender double standard exists, the allocation in which the woman contributes less and the man contributes more will be evaluated as less appropriate than the reverse scenario, where the man contributes less and the woman contributes more. This indicates that deviations from equal contributions to domestic work are evaluated asymmetrically depending on the gender of the partner who benefits from the allocation. In Section 4.1, where the model is presented, Remark 1 states the condition under which two symmetric deviations from the Egalitarian Norm are judged equally socially inappropriate. A double standard exists when this condition is not satisfied; see Definition 2.

Note that, in contrast to the framing effect discussed in Hypothesis 1, which relates to how identical allocations are perceived differently based on the gender of the proposer, the double standard focuses on whether identical deviations from an egalitarian norm are judged differently depending on the gender of the beneficiary. We are thus willing

to disentangle two cognitive biases that may contribute to the perpetuation of gendered social norms.
In our analysis, we will also assess whether age groups differ in their perceptions of this double standard. And whether the double standard is affected by the gender of the partner who proposes the allocation.
The double standard hypothesis aligns with existing literature on gender norms as a driver of household behavior; see, among others, Thébaud et al. (2021). In this context, gender norms function both descriptively and prescriptively: departing from equal contributions, people believe not only that women do more housework but also that they *should* do more. Importantly, one does not need to personally subscribe to these norms to be influenced by them. As pointed out by Ridgeway and Correll (2004), even individuals who reject gendered norms may still perceive that most others uphold them, shaping their own behavior accordingly. This perception of widespread societal expectations may reinforce gendered divisions of labor, even among those who hold progressive personal beliefs.

The last hypothesis refers to the decline of the "male breadwinner model" across generations.

Hypothesis 3. ***The decline of the "male as the breadwinner" model.*** *The traditional model, where the male partner's main sphere is the workplace while the female partner's main sphere is the household, is no longer perceived as the social norm by young adults.*

To test this hypothesis, we focus on the Part-Time vignette and examine the perceived social appropriateness of the equal-contribution scenario by dividing respondents into three age groups. In this vignette, the man works more hours and earns a higher labor income than his partner. We therefore expect that most respondents anticipate others to judge that the woman should contribute relatively more to domestic labor. However, younger respondents may be more inclined to interpret the prevailing social norm as closer to egalitarianism. Definition 3 illustrates this prediction in terms of our theoretical framework; see Section 4.2.

A progressive decline of the "male as the breadwinner" model and the rise of a "dual-earner" model has already been documented across OECD countries (Trappe et al. 2015). However, the pace and nature of this transition vary significantly across different institutional and cultural settings; see von Gleichen and Seeleib-Kaiser (2018).
In the Italian context, where adherence to traditional gender roles remains strong, this transition has been slow. This might be due to both deep-rooted cultural values and an institutional framework resistant to change. For example, Italy introduced gender-neutral parental leave only in 2000, with father participation remaining markedly low. Although childcare availability for children under three has improved—reaching a national coverage rate of 24% in 2010—regional disparities persist. In southern Italy, for instance, the enrollment rate for this age group was still less than 4% in 2010; see Del Boca et al. (2015). Despite the slow pace of the Italian transition, change is likely

to be driven by the youngest age group, consistent with previous evidence showing that younger generations tend to hold more progressive views than older cohorts. This pattern is reflected in respondents' answers across older and more recent waves of the World Values Survey (see, among many others, Fortin 2005).

Before turning to the presentation of our simple model, a few disclaimers are in order. First, throughout the paper, we refer to differences across generations or age groups. While differences across generations provide useful insights into the evolution of social norms, they should be interpreted within a broader context rather than as novel findings. In addition, we cannot claim these differences to be permanent as they could be related to differences in life stages, e.g., younger people might still have to go through parenthood, life experience, etc (see also the concluding section on this point). Third, while this study was not pre-registered, the choice of sample dimensions and treatment variations reflects our ex-ante hypotheses about the factors most likely to influence social norms. We ensured the representativeness of our sample by selecting demographic variables—such as gender, age, residence area, and education level—that we hypothesized would be critical in shaping societal views and behaviors around gender norms, particularly in household work and childcare. Likewise, our treatment variations were guided by the expectation that framing influences the formation of normative expectations.

4 A model of time allocation to household chores with gender norms

In this section, we propose a stylized model to represent the partners' situation as illustrated in the vignettes, as well as the social norm regarding contributions to household chores within a heterosexual couple.

We assume that a couple's welfare is given by the following expression:[13]

$$W = U\left(B(t_f + t_m), C_f(t_f), C_m(t_m), N^f\left(t_f^N, t_m^N, t_f, t_m\right), N^m\left(t_f^N, t_m^N, t_f, t_m\right)\right); \quad (1)$$

where t_g, with $g \in \{f, m\}$, is time devoted to household work by the partner whose gender is g, i.e., female or male. The function $B(t_f + t_m)$ denotes the benefit from a household public good which is increasing in the total time devoted to household work. The time devoted by the two partners to household chores, t_f and t_m, are thus perfectly substitutable. We let $B' > 0$, $B'' < 0$ and $B(0) = 0$.

The function $C_g(t_g)$, with $g \in \{f, m\}$, is the disutility from time spent in household

[13]We model a unitarian couple. To understand why, note that when both partners have the same working conditions (Vignette Full-Time), we do not anticipate significant differences in their bargaining weights. In contrast, when the female partner works part-time (Vignette Part-Time), assuming greater bargaining power for the male partner might be appropriate in a collective model. However, under Hypothesis 3, our focus is on measuring social approval of equal contributions. Hence, despite being in principle more appropriate, considering a collective model with bargaining weights would add unnecessary complexity in the Vignette Part-Time. Note that, in our unitary model, the proposer's identity does not directly affect the cost of the social norm. Nonetheless, we can allow γ and ρ to be proposer-specific (see below).

work by the partner whose gender is g. The function $C_g(.)$ is strictly increasing and strictly convex: $C'_g(.) > 0$, $C''_g(.) > 0$. Labor supplies and the corresponding returns (e.g., market wages) are not explicitly modelled, but differences in the shape of the functions $C_f(.)$ and $C_m(.)$ can capture possible asymmetries in the time spent in the labor market by each of the partners, as indicated in the two vignettes. Finally, $U(.)$ is a function such that: $\frac{\partial U}{\partial B} > 0$, $\frac{\partial U}{\partial C_g} < 0$, $\frac{\partial U}{\partial N^g} < 0$ and $\frac{\partial^2 U}{\partial B^2} < 0$, $\frac{\partial^2 U}{\partial C_g^2} < 0$ and $\frac{\partial^2 U}{\partial (N^g)^2} < 0$, where $g \in \{f, m\}$.

Not conforming to the social norm regarding the distribution of chores within the couple results in disutility ($\frac{\partial U}{\partial N^g} < 0$). The term N^g, with $g \in \{f, m\}$, represents the disutility generated by the social norm for each partner. Specifically:

$$N^f\left(t_f^N, t_m^N, t_f, t_m\right) = \gamma_f \max\left\{\frac{t_f^N}{t_f^N + t_m^N} - \frac{t_f}{t_f + t_m}; 0\right\} + \rho_f \max\left\{\frac{t_f}{t_f + t_m} - \frac{t_f^N}{t_f^N + t_m^N}; 0\right\};$$
$$N^m\left(t_f^N, t_m^N, t_f, t_m\right) = \gamma_m \max\left\{\frac{t_m^N}{t_f^N + t_m^N} - \frac{t_m}{t_f + t_m}; 0\right\} + \rho_m \max\left\{\frac{t_m}{t_f + t_m} - \frac{t_m^N}{t_f^N + t_m^N}; 0\right\};$$

where t_g^N and $\frac{t_g^N}{t_f^N + t_m^N}$ are the time spent in household work by an individual of gender g and the share of time spent in household work that is socially appropriate for gender g, respectively.

The functions $N^g\left(t_f^N, t_m^N, t_f, t_m\right)$ is such that the norm is "binding", i.e., it generates some disutility, when a partner's share of household work deviates from the prescribed norm. The parameter γ (respectively, ρ) measures the magnitude of the social sanction for the partner contributing less (or more, respectively) to household chores. We expect that $\gamma_g \geq \rho_g$, $\forall g \in \{f, m\}$, because society is likely to disapprove of a self-beneficial behavior more than a self-sacrificing one.[14]

When choosing the amount of time to devote to household chores, the partners take the values t_g^N, $g \in \{f, m\}$, as given. Assuming an interior solution, the first-order conditions of welfare (1) with respect to the amount of time devoted to household work by the two partners are:

$$\frac{\partial U}{\partial B}\frac{\partial B}{\partial t_g} + \frac{\partial U}{\partial C_g}\frac{dC_g}{dt_g} + \frac{\partial U}{\partial N^g}\frac{\partial N^g}{\partial t_g} = 0, \text{ with } g \in \{f, m\}; \tag{2}$$

where, because of perfect substitutability in partners' contributions to the family public good, $\frac{\partial B}{\partial t_f} = \frac{\partial B}{\partial t_m}$.

All the combinations t_f^* and t_m^* that simultaneously satisfy the system generated by the two first-order conditions defined in (2) are solutions to the welfare maximization problem.

We assume that the following three allocations, described in Vignette Part-Time and

[14] This is reminiscent of Fehr and Schmidt (2006), who examine preferences for fairness. They differentiate between envy, which arises in the context of a disadvantageous allocation, and fairness concerns, which emerge in the case of an advantageous allocation. Experimental evidence shows that the disutility from a disadvantageous allocation is greater than that from an advantageous allocation of the same magnitude.

Vignette Full-Time, satisfy the system of the two first-order conditions expressed by (2):

$$A_{V1} \equiv \left(\frac{t_f}{t_f+t_m} = \frac{1}{4}, \frac{t_m}{t_f+t_m} = \frac{3}{4}\right),$$
$$A_{V2} \equiv \left(\frac{t_f}{t_f+t_m} = \frac{1}{2}, \frac{t_m}{t_f+t_m} = \frac{1}{2}\right),$$
$$A_{V3} \equiv \left(\frac{t_f}{t_f+t_m} = \frac{3}{4}, \frac{t_m}{t_f+t_m} = \frac{1}{4}\right).$$

The survey respondents' beliefs about social approval/disapproval of the three mentioned allocations provide information on the relative sizes of the parameters γ_g and ρ_g, $g \in \{f, m\}$. Additionally, we expect that respondents' perceptions of the magnitude of the parameters γ_g and ρ_g, also depends on the gender of the partner proposing the allocation. Hence, we can add a superscript i that indicates who is proposing the allocation in the vignette:

$$\gamma_g^i \text{ and } \rho_g^i,\ g \in \{f, m\},\ i \in \{mp, wp\},$$

where mp corresponds to "man proposing" and wp corresponds to "woman proposing".

4.1 Full-time working female partner (Vignette Full-Time)

Given the symmetry between partners, a norm of equal contributions to household work is likely to exist and be expected by the respondents of the representative survey. This is confirmed by the results in Table 6 showing that the elicited social norm corresponds to the equal contribution to household chores. Let us denote the Egalitarian Norm as

$$N_E \equiv \left(\frac{t_f^N}{t_f+t_m} = \frac{1}{2}, \quad \frac{t_m^N}{t_f+t_m} = \frac{1}{2}\right).$$

This egalitarian norm generates social disapproval when partners do not contribute equally to the public good.

When $t_m < \frac{1}{2} < t_f$, the male partner experiences disutility, denoted by $\gamma_m \left(\frac{1}{2} - \frac{t_m}{t_f+t_m}\right)$, because he deviates from the egalitarian norm with an advantageous allocation of time. Simultaneously, the female partner suffers disutility, represented by $\rho_f \left(\frac{t_f}{t_f+t_m} - \frac{1}{2}\right)$, as she deviates from the egalitarian norm with a disadvantageous allocation of time. The parameters γ_m and ρ_f indicate the strength of social disapproval and, consequently, the cost of deviating from the social norm.

If $t_m > \frac{1}{2} > t_f$, the opposite situation occurs. The male partner experiences disutility $\rho_m(\frac{t_m}{t_f+t_m} - \frac{1}{2})$ and the female partner suffers disutility $\gamma_f \left(\frac{1}{2} - \frac{t_f}{t_f+t_m}\right)$.

Let us consider the three possible allocations.

- In A_{V2}, where $\frac{t_f}{t_f+t_m} = \frac{t_f^N}{t_f^N+t_m^N} = \frac{t_m}{t_f+t_m} = \frac{t_m^N}{t_f^N+t_m^N} = \frac{1}{2}$, partners adhere to the norm and, thus, do not experience disutility.

- In A_{V1},where $\frac{t_f}{t_f+t_m} = \frac{3}{4} > \frac{t_m}{t_f+t_m} = \frac{1}{4}$, the norm is binding for both partners. One can check that $\gamma_m \left(\frac{1}{2} - \frac{t_m}{t_f+t_m}\right) = \frac{1}{4}\gamma_m$ and $\rho_f \left(\frac{t_f}{t_f+t_m} - \frac{1}{2}\right) = \frac{1}{4}\rho_f$. Hence, the overall disutility from deviations from the norm in allocation A_{V1} is $\frac{1}{4}(\gamma_m + \rho_f)$.

- In A_{V3}, where $\frac{t_f}{t_f+t_m} = \frac{1}{4} < \frac{t_m}{t_f+t_m} = \frac{3}{4}$, the disutilities from deviating from the norm are $\frac{1}{4}\gamma_f$ and $\frac{1}{4}\rho_m$, respectively. Thus, in allocation A_{V3}, overall disutility from deviations from the norm is $\frac{1}{4}(\gamma_f + \rho_m)$.

It follows from the reasoning above that:

Remark 1. *Allocations A_{V1} and A_{V3}, representing two symmetric deviations from the Egalitarian Norm, will be judged equally socially inappropriate if and only if $\gamma_m + \rho_f = \gamma_f + \rho_m$.*

We are now ready to state how the two hypotheses based on Vignette Full-Time can be interpreted using the theoretical model.

Definition 1. ***Framing effect.*** *The self-beneficial allocation is evaluated as less appropriate when proposed by the woman than when proposed by the man: $\gamma_f^{wp} \geq \gamma_m^{mp}$. This implies that allocation A_{V3} (where the woman proposes that she contributes less) is rated as less appropriate than its mirror image allocation A_{V1} (where the man proposes that he contributes less).*

This illustrates our Hypothesis 1. Note that we are agnostic as to the parameter ρ_g^i—about the *self-detrimental* allocation—and do not hold any specific expectation on whether $\rho_m^{mp} \gtrless \rho_f^{wp}$.

Using Remark 1, our Hypothesis 2 can be defined as follows.

Definition 2. ***Gender Double Standard****. The two deviations from the egalitarian norm (contributing more or contributing less) are judged differently: $\gamma_m + \rho_f \neq \gamma_f + \rho_m$. In particular, we expect that $\gamma_m + \rho_f < \gamma_f + \rho_m$, meaning that the allocation A_{V1}—woman contributing more and man contributing less—will be rated as more appropriate (i.e. less negative) than its mirror image allocation A_{V3}—woman contributing less and man contributing more.*

4.2 Part-time working female partner (Vignette Part-Time)

Here, the disutility from time spent in household work is higher for the male partner. Hence, it is plausible to assume that the social norm is now such that the woman devotes more time than her partner to domestic work:

$$\frac{t_f^N}{t_f^N + t_m^N} > \frac{1}{2} > \frac{t_m^N}{t_f^N + t_m^N} \Rightarrow \left(t_f^N - t_m^N\right) > 0. \tag{3}$$

Note that the closer $\frac{t_f^N}{t_f^N+t_m^N}$ and $\frac{t_m^N}{t_f^N+t_m^N}$ are to $\frac{1}{2}$ (and $t_f^N - t_m^N$ to zero), the closer society is to the Egalitarian Norm.

Let us now consider the allocation entailing equality of contributions:
$A_{V2} = \left(\frac{t_f}{t_f+t_m} = \frac{1}{2}, \frac{t_m}{t_f+t_m} = \frac{1}{2}\right)$. Under (3), the total disutility from norm deviation

generated by such allocation is:

$$\gamma_f \left(\frac{t_f^N}{t_f^N + t_m^N} - \frac{1}{2} \right) + \rho_m \left(\frac{1}{2} - \frac{t_m^N}{t_f^N + t_m^N} \right); \tag{4}$$

where the female partner is deviating from the norm because she does not contribute enough, while the male partner contributes too much. Intuitively, the perceived total disutility expressed in (4) is inversely related to the perceived social appropriateness of allocation A_{V2}.

Definition 3. ***Decline of the "male as the breadwinner" model.*** *Specifically, the respondents' perception of the difference* $\left(t_f^N - t_m^N\right)$ *in Vignette Part-Time is age-specific. The perceived appropriateness of the equal share of family chores described in allocation* A_{V2} *is higher for young adults:*

$$\left(t_f^N - t_m^N\right)_{25\text{-}34} < \left(t_f^N - t_m^N\right)_{35\text{-}49}, \left(t_f^N - t_m^N\right)_{50\text{-}64}.$$

5 Results

For each of our three hypotheses, we illustrate our findings first at the aggregated and, second, at the individual level. Finally, we highlight the correlation between social norms elicited in our sample and the outcome of female labor market outcomes measured by administrative data.

To aggregate individual answers, following a common procedure in the experimental literature (see, e.g. KW and Barr et al. 2018)[15] we use the *appropriateness norm rating* obtained by converting subjects' answers to numerical values. Specifically, we attribute to every Likert scale item a numerical counterpart: Very Appropriate is converted to the value $+1$, Somewhat Appropriate to $+0.33$, Somewhat Inappropriate to -0.33, finally, Very Inappropriate is converted to -1. In this way, we represent Likert scale items as evenly spaced, this allows us to perform parametric tests but imposes an additional assumption on our data. To take into account this assumption, we replicate our tests using non-parametric tests that do not impose evenly spacing on our Likert scale items.[16]

In all our analyses, we use sample weights, which guarantee the representativeness of our sample for within/between-group comparison and regression analysis.[17] Whenever we use statistical tests, we follow the Benjamini-Hochberg False Discovery Rate method (Benjamini and Hochberg 1995) for multiple test adjustment: we sort the p-values in ascending rank and multiply each by the number of separate tests being performed before dividing each by its rank- thus greater adjustments are made to smaller p-values. Table 6 and Table 7 present the distribution of answers for Vignette Full-Time and

[15] The same transformation was used among others by Chang et al. (2019), d'Adda et al. (2016), Erkut et al. (2015), Gächter et al. (2017), Gächter et al. (2013), Kimbrough and Vostroknutov (2016), Schneeberger and Krupka (2021) and Veselỳ (2015).

[16] When a test has been replicated using non-parametric tests, we report whether the result holds, or not.

[17] Specifically, we implement the command "svy" in Stata.

Vignette Part-Time, respectively; the social norm for each of the three scenarios[18] is enclosed in a rectangle, "strong" norms (i.e. norms that are shared by the majority of our respondents) are in boldface.

	Vignette Full-Time		
	Woman contributes less	Equal contribution	Man contributes less
Very Inappropriate	49.34	1.83	32.44
Somewhat Inappropriate	32.55	6.27	34.12
Somewhat Appropriate	14.44	24.75	25.88
Very Appropriate	3.67	**67.15**	7.56
Mean Rating	-.5164	.7142	-.2761

Table 6: **Family Norm, Vignette Full-Time**

Vignette Full-Time: "Imagine Antonio and Francesca: they are married or cohabiting. They both work the same number of hours, earn roughly the same amount of money, and have the same career trajectories. They have no children and no one to help them with the housework." The elicited social norms are enclosed in a rectangle; strong norms (i.e., norms shared by the majority of the sample) are presented in boldface.

	Vignette Part-Time		
	Woman contributes less	Equal contribution	Man contributes less
Very Inappropriate	**57.38**	11	10.03
Somewhat Inappropriate	25.63	33.73	17.47
Somewhat Appropriate	13.37	34.49	40.98
Very Appropriate	3.62	20.79	31.52
Mean Rating	-.5782	.1004	.2925

Table 7: **Family Norm, Vignette Part-Time**

Vignette Part-Time: "Imagine Giulio and Silvia: they are married or cohabiting. Giulio works twice as many hours as Silvia and earns about twice as much. They have no children and no one to help them with the housework." The elicited social norms are enclosed in a rectangle; strong norms (i.e., norms shared by the majority of the sample) are presented in boldface.

5.1 Framing and Gender Double Standard

In this section we focus on Vignette Full-Time which depicts a set-up where the partners share the same working arrangements: they work the same number of hours per week and earn the same amount of money.

At the aggregate level, Table 6 shows the distribution of answers for the three scenarios. For each scenario, the norm corresponds to the modal response. The most appropriate behavior when partners have the same working condition is the "equal contribution to household chores," rated as very appropriate by 67.15% of respondents. Any deviation from equal contribution is seen as socially inappropriate. However, as we show below, the perceived inappropriateness of this deviation depends on both the

[18] In what follows, we always refer to "woman contributes less", "equal contribution", and "man contributes less" as our three scenarios.

gender of the partner proposing the allocation (framing effect) and the gender of the partner contributing less (double standard).

5.1.1 Framing effects

Studying the framing effect, we focus on our treatment variable, namely the gender of the partner proposing the allocation of household chores in the three scenarios. According to our Hypothesis 1, we examine whether a woman proposing an allocation that benefits[19] her (and disadvantages her partner) is rated as more socially inappropriate than a man proposing an allocation that benefits him. This allows us to explore whether normative societal expectations differ based on the gender of the proposer in scenarios of unequal contribution.

Panel a) Woman Proposing			
	Woman Contributes Less	Equal Contribution	Man Contributes Less
Very Inappropriate	**60.06**	1.31	24.21
Somewhat Inappropriate	27.93	4.92	36.31
Somewhat Appropriate	9.53	25.48	29.85
Very Appropriate	2.48	**68.29**	9.63
Mean Rating	-.6365	.7376	-.1671
Panel b) Man Proposing			
Very Inappropriate	36.61	2.43	42.21
Somewhat Inappropriate	38.04	7.87	31.52
Somewhat Appropriate	20.28	23.90	21.18
Very Appropriate	5.08	**65.80**	5.09
Mean Rating	-.374	.6866	-.4054
Panel c) Mean Differences (p-value), framing			
	Proposer's Advantage	Equality	Recipient's Advantage
	-.2311 (.0000)	.0510 (.0630)	.2069 (.0000)
Panel d) Mean Differences (p-value), woman versus man contributes less			
Woman Proposing	-.4695 (.0000)	Man Proposing	.0314 (.3855)

Table 8: Family Norm by proposer's gender, Vignette B

Table 8 presents the distribution of the answers for Vignette Full-Time in the three scenarios, distinguishing between "woman proposing" and "man proposing" (our treatments) in panels a) and b), respectively. In both treatments, proposing a self-beneficial allocation is associated with higher judgments of social inappropriateness. Specifically, "very inappropriate" represents the social norm when the partner proposing the allocation contributes less. Nonetheless, response patterns differ by gender: 60% of respondents evaluate a woman making such a proposal as very inappropriate, whereas the corresponding share is 42% when the proposer is a man.

[19] In what follows, when we talk about allocations benefiting a partner, we refer to allocations that result in a lower share of house chores for that partner, this in turn implies that the other partner will enjoy an allocation that disadvantage him/her, i.e., such that he/she has a higher share of house chores.

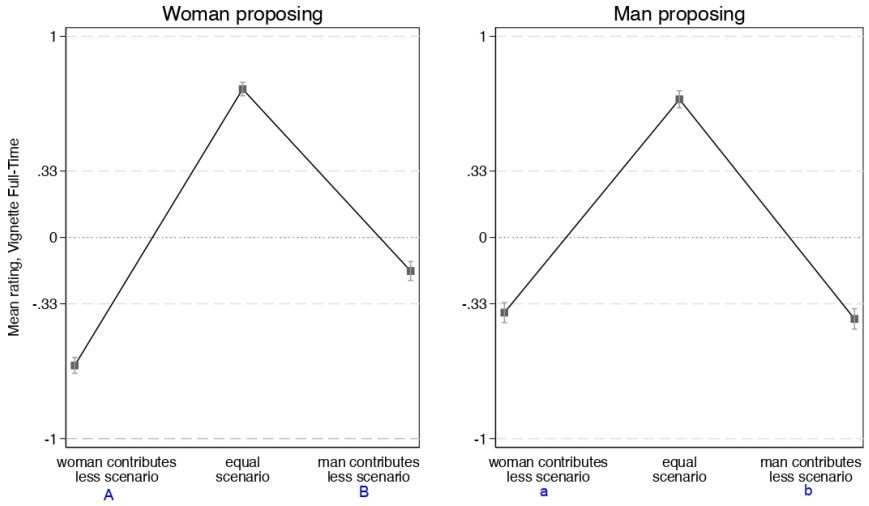

Figure 1: **Norm function for Vignette Full-Time:** "Imagine Antonio and Francesca: they are married or cohabiting. They both work the same number of hours, earn roughly the same amount of money, and have the same career trajectories. They have no children and no one to help them with the housework." 95% Confidence intervals are displayed.

Figure 1 plots mean ratings reported in Table 8 for the two treatments. Note that, for checking the framing effect in Figure 1, one must compare allocation A with allocation b, and B with a.

Table 9 is similar to Table 8, but extends the analysis by distinguishing across age groups. It confirms that the elicited norm for the equal contribution scenario is Very Appropriate across all age groups, regardless of the proposer's gender. Proposing a self-advantageous allocation is consistently rated as Very Inappropriate, but this evaluation is more prevalent when the proposer is the female partner, and less so when the proposer is male (see the first three columns on the left).

Looking at the young and middle generations, Table 9 shows that the allocation where the woman contributes less is rated by the majority of respondents as Very Inappropriate when the proposer is a woman, while it is considered Somewhat Inappropriate when the proposer is the male partner

Figure A1, in the Appendix, presents the same results contained in Table 9 plotting the norm function by age group.

To formally test our Hypothesis 1, in panel c) of Table 9 we test the null hypotheses of equality of means between woman and man proposing, in the three scenarios considered by the Vignette Full-Time: (i) proposer's advantage, i.e. the proposer is contributing less; (ii) equal contribution; (iii) recipient's advantage, i.e. the recipient contributes less than the proposer. For each scenario, panel c) of Table 9 reports the mean differences,[20]

[20]When we talk about mean differences, we refer to differences between mean ratings. In this case, the mean differences refer, for example, to the difference between the mean rating for the proposer's advantage scenario in the age group 25–34 woman proposing, and the mean rating for the proposer's advantage scenario in the age group 25–34 man proposing.

and the adjusted p-valued in parenthesis. Results are coherent with our Hypothesis 1, as we do find significant differences in the mean ratings for cases (i) and (iii) for the middle and elder age groups. That is, we find that the two elder age groups exhibit a framing effect: the proposer contributing less (more) is rated differently based on their gender.

Panel a) Woman Proposing

	Woman Contributes Less			Equal Contribution			Man Contributes Less		
	25-34	35-49	50-64	25-34	35-49	50-64	25-34	35-49	50-64
Very Inappropriate	**56.84**	**58.15**	**63.62**	.95	1.56	1.24	25.51	22.19	25.65
Somewhat Inappropriate	30.5	28.17	26.41	5.43	4.72	4.88	40.52	33.85	36.8
Somewhat Appropriate	9.12	10.94	8.27	23.23	28.19	23.77	26.3	32.91	28.42
Very Appropriate	3.54	2.74	1.7	**70.39**	**65.53**	**70.11**	7.68	11.05	9.12
Mean Rating	-.6036	-.6109	-.6791	.7531	.7171	.751	-.2252	-.1145	-.1929

Panel b) Man Proposing

	Man Contributes Less			Equal Contribution			Woman Contributes Less		
Very Inappropriate	49.9	37.21	42.91	1.7	3.22	2.08	32.96	32.81	41.92
Somewhat Inappropriate	28.46	33.07	31.64	5.49	8.64	8.36	37.26	41.77	35.02
Somewhat Appropriate	17.19	23.29	21.25	23.34	26.23	22.06	24.4	19.68	18.74
Very Appropriate	4.45	6.42	4.2	**69.48**	**61.91**	**67.49**	5.38	5.74	4.32
Mean Rating	-.4917	-.3402	-.4214	.7367	.645	.6993	-.3182	-.3436	-.4297

Panel c) Framing: Mean differences (p-value)

	Proposer's Advantage		Equality		Recipient's Advantage	
Age group: 25–34	-.1119	(.1827)	.0164	(.7825)	.0930	(.2780)
Age group: 35–49	-.2707	(.0000)	.0721	(.1079)	.2291	(.0000)
Age group: 50–64	-.2577	(.0000)	.0517	(.3180)	.2368	(.0003)

Panel d) Gender double standard: Mean differences (p-value), woman versus man contributes less

	age group: 25–34		age group: 35–49		age group: 50–64	
Woman Proposing	-.3784	(.0000)	-.4965	(.0000)	-.4861	(.0000)
Man Proposing	.1734	(.0351)	-.0035	(1.0000)	-.0083	(.8961)

Table 9: **Family Norm by proposer's gender and age groups, Vignette Full-Time**

Vignette Full-Time: "Imagine Antonio and Francesca: they are married or cohabiting. They both work the same number of hours, earn roughly the same amount of money, and have the same career trajectories. They have no children and no one to help them with the housework."
Panel c) Benjamini-Hochberg adjusted p-values in parenthesis refer to a test of equality within age groups. Results are replicated with Wilcoxon rank-sum test. Panel d) Benjamini-Hochberg adjusted p-values in parenthesis refer to a test of equality of woman and man contributes less scenario within age groups. Results are replicated with Wilcoxon rank-sum test.

These findings can be summarized as follow:

Result 1. Framing: *Considering full-time dual-earner couples, framing effects are documented for mid-lifers and seniors but not young adults. In the two elder groups, a woman proposing a self-benefiting chores distribution is perceived to be less socially appropriate than a man proposing a self-benefiting chore distribution, while a man proposing self-sacrificing arrangements is perceived as less socially appropriate than a woman proposing a self-sacrificing arrangement. This disparity disappears among young adults.*

The finding that a woman proposing a chore distribution favorable to herself yet unfavorable to her partner is perceived as less appropriate than a man doing the same thing can be attributed to entrenched gender roles. Traditionally, domestic responsibil-

ities are viewed as the woman's domain; thus, when a woman attempts to assign more household tasks to her partner, it defies these stereotypes and invites greater societal sanction. Conversely, the observation that a man suggesting a chores allocation that is less favorable for himself, but beneficial to his partner faces more negative judgment than a woman reflecting traditional notions of masculinity. A man assuming primary responsibility for domestic chores challenges conventional masculine roles, leading to societal disapproval. Among young adults, however, such counter-stereotypical behaviors do not seem to be sanctioned, possibly indicating a positive shift towards gender-neutral and egalitarian attitudes in managing household responsibilities within Italian society.

5.1.2 Double standard

Recall that, unlike framing—where the relevant factor is the proposer's gender—the double-standard focuses on the gender of the beneficiary of the unequal allocation. In other words, here, we compare the two mirror-image deviations from equal contribution—"woman contributes less" versus "man contributes less," initially disregarding who makes the proposal. We will return to this aspect in the discussion that follows.

First, Table 6 reveals a marked difference in mean ratings between the scenarios "woman contributes less" and "man contributes less." When looking at the modal response, "woman contributes less" is classified as *very inappropriate*, whereas its mirror allocation favoring men is deemed only *somewhat inappropriate*. This disparity suggests a double standard: identical deviations from the equal contribution are evaluated differently depending on the gender of the person contributing less.

To analyze the double standard by treatment we can use again Figure 1 and Table 8. As mentioned above, Figure 1 displays the norm function for the Full-Time Vignette across the overall sample by treatment, while Table 8 reports the elicited norms along with statistical tests. For checking the double standard in Figure 1, one must compare allocation A with allocation B, and a with b. Specifically, the analysis contrasts the perceived appropriateness of self-beneficial versus self-detrimental allocations (by proposer gender).

Notably, Figure 1 shows that the double standard arises exclusively in the "woman proposing" treatment. In words, when the woman proposes the chores allocation, the woman contributing less/man contributing more scenario is perceived as less socially appropriate than the mirror image scenario of woman contributing less/man contributing more. Conversely, the two deviations from equal contributions are evaluated in the same way when the man proposes the deviation. This analysis shows that the double standard bias is primarily driven by the treatment "woman proposing."

Looking again at Figure A1 in the Appendix, we observe that the younger generation is affected by the double standard, albeit to a lesser extent.

To sum up, our analysis shows that Hypothesis 2 holds only in the "woman proposing" treatment. In terms of our theoretical model, all this implies that $\gamma_m^{wp} + \rho_f^{wp} < \gamma_f^{wp} + \rho_m^{wp}$, but $\gamma_m^{mp} + \rho_f^{mp} \sim \gamma_f^{mp} + \rho_m^{mp}$. This highlights the interplay between framing effects and

double standards.

We next explore the presence of the double standard bias by analyzing responses at the individual level. Table 10 presents the average marginal effects for the logistic regression estimating the probability of identifying a norm that rates the "woman contributes less" scenario as less socially appropriate than the "man contributes less" scenario in the overall sample (to ease comparison, we report the mean of the dependent variable: 0.397). Hence, we check here the probability that respondents rate $\gamma_m + \rho_f < \gamma_f + \rho_m$, irrespective of the proposer's gender. For a description of all the variables used in our analysis see Table A3 in the Appendix. In model (1) we control for our reference group categories, together with a dummy for the framing. In model (2) we add controls for the respondent having relocated to a different geographical area; for example, the category 'Moved North' identifies respondents who are resident in a macro-area that is northern than the one in which the respondent was born. We lost 58 observations as we either did not have reliable information on the macro-area of birth or because of a foreign place of birth. In model (3), we add controls for civil status and the respondent's parenthood. Model (4) adds controls on education and job status. Table OA3 in the Appendix presents additional models controlling for a set of personality traits (model (5)) and a set of controls at the municipality level (model (6)) using data from the Urban Index (https://www.urbanindex.it).

Both in Tables 10 and OA3 we find some evidence of a difference in this probability for the young generation in models (1) and (2), but the coefficient is no longer significant once we add controls. The sign of the coefficient for young adults is, however, negative. This suggests that young adults are less likely to exhibit a gender double standard, but we no longer capture this once we add controls on family formation.
Not surprisingly, in all models analyzed, we find a statistically significant effect of the gender of the proposer of the housework chores allocation: when the allocation is proposed by the woman, the probability of perceiving a norm of higher inappropriateness for her (with respect to the male partner) is increased by about 27pp. This confirms the pattern observed in Figure 1, indicating that framing and double standard biases are interconnected.

In what follows, we briefly describe results from other regressions carried out to better understand the gender double standard and its link with framing. In the online Appendix, Table OA4 replicates the estimates contained in Table 10, including interactions between gender, geographical areas, proposer's gender, and age groups. Panel a) model (1) includes interactions between age groups and gender, model (2) includes interactions between age groups and geographical areas, and model (3) includes interactions between age, gender, and geographical area. All specifications include a control for the gender of the proposer. We find suggestive evidence that the estimates for the younger generation are driven by the males in the South and Islands, while the effects for the middle-aged generation seem to be driven mostly by the males in the center. Panel b) includes interactions between the gender of the proposer and the gender of the respondent. We would like to stress that these regressions provide only suggestive

evidence, and are intended to try to cast a light on determinants of differences in elicited norms between generations.

	(1)	(2)	(3)	(4)
Dependent Variable	1 if identifies a norm stigmatizing the "Woman contributes less" more than the "Man contributes less" scenario, 0 otherwise			
Female	-0.052	-0.035	-0.041	-0.042
	(0.0275)	(0.0280)	(0.0281)	(0.0304)
Age Groups (Baseline: 50-64)				
25-34	-0.078*	-0.078*	-0.049	-0.040
	(0.0376)	(0.0381)	(0.0416)	(0.0418)
35-49	0.024	0.021	0.026	0.033
	(0.0311)	(0.0317)	(0.0320)	(0.0318)
Geographical Areas (Baseline: South and Islands)				
North	-0.021	-0.018	-0.016	-0.009
	(0.0310)	(0.0331)	(0.0329)	(0.0332)
Centre	-0.020	-0.038	-0.033	-0.028
	(0.0399)	(0.0418)	(0.0417)	(0.0419)
Relocated to a different Geographical Area (Baseline: Did not move)				
Moved North		0.038	0.038	0.042
		(0.0423)	(0.0429)	(0.0427)
Moved South		0.072	0.064	0.061
		(0.0713)	(0.0704)	(0.0704)
Civil Status (Baseline: Single, Widower, Separated-Divorced)				
Married or Cohabitant			0.049	0.052
			(0.0344)	(0.0343)
Having Children			0.043	0.039
			(0.0336)	(0.0336)
Framing: Woman Proposing	0.276***	0.277***	0.277***	0.277***
	(0.0273)	(0.0277)	(0.0276)	(0.0275)
Controls				
Education/Job	-	-	-	✓
Observations	1501	1443	1443	1443

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: **Gender Double Standard, Vignette Full-Time**

Average marginal effects for the probability of rating the woman contributing less scenario less appropriate than the man contributing less scenario. In columns (2)-(4) we loose data on 58 observations as we do not have reliable information on the geographical area at birth, or as respondents were born abroad.

Again in the appendix, Table A7 and Table A8 replicate Table 10, disaggregating by our treatments. Thus, they present the average marginal effects for the logistic regression estimating the probability of identifying a norm that rates the "woman contributes less" scenario as less socially appropriate than the "man contributes less" scenario, respectively in the woman proposing (Table A7) and man proposing (Table A8) treatments. Thus, we are interested in possible drivers of the gender double standard. The set of independent variables is the same as used in Table 10. In Table A7 we are unable to find statistically significant effects, yet it is worth noting that being in the younger age group is associated with a lower (yet not statistically significant) probability of rating a woman proposing to do less as less appropriate than a woman proposing to do more,

this difference (approximately 4pp, where the mean of our dependent variable is .526) switches sign once we add additional controls on family formation. In Table A8 we find that the younger generation is associated with a lower probability of identifying a norm that punishes the man offering to do more, more than the man offering to do less, when he proposes the chores allocation (approximately 11pp, where the mean of our dependent variable is .245). This is in line with our results from Table 9, and suggests a possible shift in the younger generation's attitudes toward more egalitarian gender norms.
Hereafter, we summarize the results of the gender double standard.

Result 2. Gender double standard: *In the context of a full-time dual-earner couple, a woman contributing less than her partner is perceived as less socially appropriate than a man in a similar situation. However, this result is driven by the woman-proposing treatment. When the man proposes the chore allocation, deviations from equal contribution are rated equally.*

Despite the prevalence of an egalitarian norm for dual-earner couples with similar working conditions, respondents display a societal bias that penalizes women more than men for contributing less to domestic chores under identical circumstances. This gender-based double standard is primarily driven by the "woman proposing" treatment. When splitting the sample by treatment, we find that a deviation from equal contribution in favor of the woman is judged as less socially appropriate when she proposes it, whereas the same deviation is not penalized when proposed by the man.
More generally, across all age groups, we observe that a woman suggesting to do less is judged more harshly than when she suggests doing more, while a man's proposal to do less is rated similarly to one in which he proposes to do more. This asymmetry reflects the persistence of traditional expectations—such as the "woman as homemaker" view—that assign women greater responsibility for domestic labor. Notably, younger respondents appear less influenced by this bias than older cohorts, suggesting a generational shift away from these traditional norms.

5.2 Decline of the "Male as the Breadwinner" Model

Vignette Part-Time describes the set-up in which the partners are ex-ante unequal: the male partner works and earns about twice as much as the female partner depicting the traditional "male as the breadwinner model". Table 7 presents the distribution of answers to Vignette Part-Time at the aggregate level as well as the mean rating for each scenario. When the woman contributes less, most respondents expect other group members to judge this scenario as Very Inappropriate. Interestingly, both the scenario where partners contribute equally and the scenario where the man contributes less than the woman are judged as Somewhat Appropriate by the majority of respondents. However, there is a slight difference in the percentages: 34.49% of respondents rates that equal

contribution as Somewhat Appropriate, while a higher 40.98% rates it in the same way when the man contributes less. Moreover, when comparing the mean ratings between these two scenarios, the difference is statistically significant (mean difference: -.1921; t-test for the equality of means p-value = .0000).[21] Note that these answers suggest that more effort exerted in household chores by the woman could, in principle, compensate for the larger effort exerted in the labor market by the man.
To test our third hypothesis predicting a decline of the "male as the breadwinner" model among young adults, we disaggregate answers by age groups.

Figure 2 presents the mean rating for Vignette Part-Time and the three different scenarios evaluated distinguishing between the three age groups. The three scenarios display an appropriateness rating decreasing in the age groups.
Table A6, in the Appendix, presents the elicited norms for different age groups. According to Hypothesis 3 we elicit a norm that is more in line with the "male as breadwinner model" for the elder generation (compared to the middle-aged and young adults). In particular, we find that for the scenario "man contributes less" the elicited norm is *Somewhat Appropriate* for all generations. For the scenario "equal contribution", the elicited norm differs across generations: for the elder generation, it is *Somewhat Inappropriate* while for the other two generations, it is *Somewhat Appropriate*.
We next test this by performing t-tests for the equality of means for each scenario, between age groups. For example, column 1 compares age groups 25–34 vs 35–49 and presents the difference in the mean ratings for the "woman contributes less" scenario between the two age groups, and reports the p-value associated with a test of equality of means in parentheses.
We find a statistically significant difference in the "equal contribution" scenario between the younger and the two elder generations, which confirms our third hypothesis.

To dig deeper into the determinants of respondents' perceived norms, we present the average marginal effects for a logistic model in Table 11. We estimated the probability that respondents evaluate the equal contribution scenario in Vignette Part-Time as either *Very Appropriate* or *Somewhat Appropriate*. This reflects the likelihood of perceiving the gender norm regarding household chores as egalitarian, even when the male partner contributes more to the labor market. To ease comparison, we report that the mean of our dependent variable in model (1) is .553. In Table 11, we include the same controls as in Table 10.

Table 11 shows that being a young adult or mid-lifer is associated with a positive and significant increase in the probability of perceiving the gender norm as egalitarian, compared to the older age group. Specifically, the probability increases by approximately 13 to 15 percentage points for young adults and around 9 percentage points for mid-lifers. All other controls, including geographical areas of living, are not significant.[22]

[21] This result is replicated with the Wilcoxon rank-sum test.

[22] Table OA5 in the Online Appendix presents the full set of estimates. We find a positive association between reporting "work" as the most important trait in life, and strongly disagreeing with the claim "A woman should be ready to reduce the time devoted to her job for family reasons." Finally, we find a negative association with the trait conscientiousness.

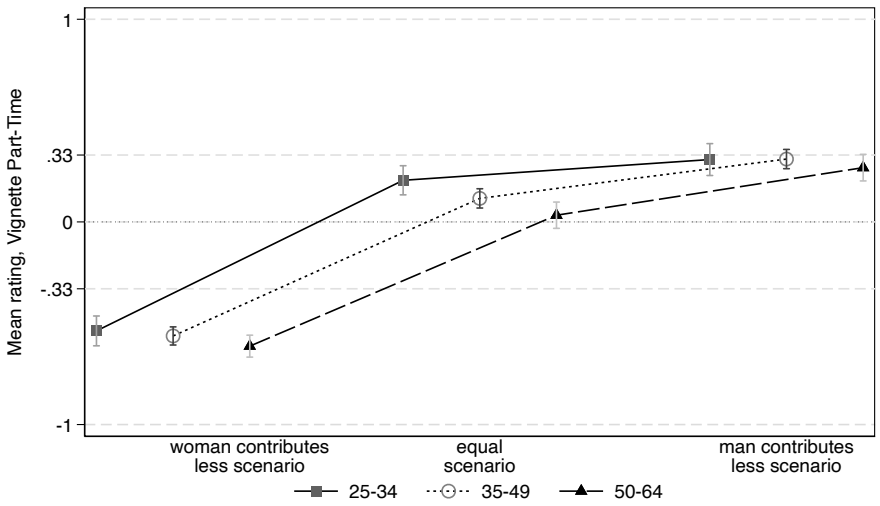

Figure 2: **Norm function for Vignette Part-Time**: "Imagine Giulio and Silvia: they are married or cohabiting. Giulio works twice as many hours as Silvia and earns about twice as much. They have no children and no one to help them with the housework."
95% Confidence Intervals are shown.

We replicated the estimates contained in Table 11 including interactions between gender, geographical areas, proposer's gender, and age groups. Table OA6 in the Online Appendix presents the results for those interactions. Specifically, Panel a) model (1) includes interactions between age groups and gender, model (2) includes interactions between age groups and geographical areas, and model (3) includes interactions between age, gender, and geographical area. All specifications include a control for the gender of the proposer. We do not find evidence that a specific group is driving the estimates for the younger generation, while the effects for the mid-lifers seem to be driven mostly by the North. Finally, Panel b) presents the interaction between the gender of the proposer and the gender of the respondent.
The main results from this section are summarized below.

Result 3, Decline of the breadwinner model: *When the male partner works and earns twice as much as the female partner, the probability of perceiving a norm of appropriateness for the equal share of family chores decreases monotonically in the age groups.*

We interpret this result as the "decline of the man as the breadwinner model" in favor of the "dual-earner model". Younger generations appear to endorse a more progressive norm, where *progressiveness* is understood as a more equitable distribution of household chores within the couple. In this context, it is important to clarify that progressiveness does not necessarily imply an equal split of tasks, but rather a fairer allocation of domestic responsibilities.

Model	(1)	(2)	(3)	(4)
Dependent Variable	1 if identifies Very or Somewhat Appropriate as norm in the equal share scenario, 0 otherwise			
Independent Variables				
Female	-0.019	-0.018	-0.015	-0.025
	(0.0291)	(0.0296)	(0.0297)	(0.0318)
Age Groups (Baseline: 50-64)				
25-34	0.143***	0.145***	0.131**	0.126**
	(0.0404)	(0.0409)	(0.0436)	(0.0439)
35-49	0.085**	0.092**	0.088**	0.087*
	(0.0327)	(0.0332)	(0.0337)	(0.0338)
Geographical Area of Residence (Baseline: South and Islands)				
North	0.055	0.038	0.036	0.040
	(0.0329)	(0.0350)	(0.0349)	(0.0356)
Centre	0.009	-0.016	-0.018	-0.016
	(0.0424)	(0.0436)	(0.0436)	(0.0440)
Relocation to a different Geographical Area (Baseline: did not move)				
Moved North		0.044	0.040	0.038
		(0.0452)	(0.0453)	(0.0451)
Moved South		-0.063	-0.059	-0.059
		(0.0748)	(0.0756)	(0.0755)
Civil Status (Baseline: Single, Widower, Separated-Divorced)				
Married or Cohabitant			-0.002	-0.001
			(0.0368)	(0.0367)
Having Children			-0.031	-0.028
			(0.0353)	(0.0352)
Framing: Woman Proposing	0.009	0.003	0.003	0.005
	(0.0291)	(0.0296)	(0.0296)	(0.0297)
Controls				
Education & Job Status	-	-	-	✓
Observations	1501	1443	1443	1443

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: **Decline of the breadwinner model, Vignette Part-Time**

Average marginal effects for the probability of rating the equality scenario Very Appropriate or Somewhat Appropriate. In columns (2)-(4) we loose 58 observations as we do not have reliable information on the geographical area of birth, or as respondents were born abroad.

To what extent is the younger generation different from the two elder generations? To answer this question, we performed additional analyses. When respondents are asked to assign points to different dimensions of life based on their perceived importance, the results suggest that generations are indeed different.[23] Specifically, we find that the younger generation assigns more importance to its professional career compared to the other age groups (Benjamini-Hochberg adjusted p-values for t-tests on the number of points assigned to the work dimension: age group 25-34 vs age group 35-49: difference=.78 p-value=.535; age group 25-34 vs age group 50-64: difference=5.11 p-value=.0005; age group 35-49 vs age group 50-64: difference=4.33 p-value=.0002).

To understand whether we are capturing a real change in the social norm, we investigate further.

[23]The question asks "Assign a total of 100 points to indicate the degree of importance you currently give to these areas of your life." The areas, presented in random order, are the following: a) My free time (e.g., hobbies, sports, recreational activities, and socializing with friends); b) My community (e.g., volunteer, union, and political organizations); c) My work, d) My religion (religious activities and beliefs); e) My family.

Before considering the first analysis, recall that we are eliciting *perceived* social norms. One might think that different elicited norms may derive from different probabilities of guessing beliefs about others' beliefs correctly across generations. Therefore, we examine the probability of correctly identifying the social norm within the reference group to check whether it is affected by age. Results from this exercise are presented in Table OA7 in the Online Appendix. We define a dummy variable that equals 1 if the individual correctly identifies the response most commonly given by his/her reference group (i.e., correctly guesses the social norm) and 0 otherwise. Findings from these regressions indicate that age does not predict the probability of correctly perceiving the social norm. Therefore, we can rule out that the observed responses from younger participants are due to a greater ability to identify higher-order beliefs compared to the elder age groups. In other words, we find no support for the idea that our estimates are influenced by participants' ability to correctly perceive the norm.

Second, we examine participants' personal opinions on the same vignette, expressed after the incentivized procedure, and relate these responses to their views on perceived social norms. To this end, we replicate the analysis in Table A6 and the regressions in Table 11 using personal opinions instead of perceived norms. Specifically, we estimate a model for the probability of personally rating the equal contribution scenario as Very Appropriate or Somewhat Appropriate. Results from this analysis are reported in the Online Appendix, in Tables OA8 and OA9, respectively.

Table OA8 shows distributions and mean ratings by age groups that are more similar to each other compared to those displayed in Table A6. Table OA9 documents that the coefficients associated with the age groups do not achieve statistical significance. Thus, we do not find evidence that the senior generation holds more traditional *personal opinions* compared to the younger age groups.

Together, evidence from this section suggests that what we are documenting is a genuine shift in social norms among young Italians, moving away from the male breadwinner model towards a more egalitarian view of the couple.

5.3 Higher-Order Beliefs and Labor Market Outcomes

In this section, we explore the association between our measures of social norms and women's labor market outcomes in Italy. We focus on the "equal contribution" scenario from Vignette Part-Time and exploit administrative data on female labor market outcomes from the Italian Institute of Statistics (ISTAT). This analysis is inspired by Fortin (2005), who examined how country-specific agreement with certain statements from the World Values Survey (WVS)—used as proxies for social norms—correlates with differences in women's labor market outcomes across countries. Similarly, we explore how age and geographical variations in perceived norms in Italy correlate with differences in female employment rate for Italian women aged 20-64.[24] We use publicly available data

[24] Female employment rate is defined as the percentage of employed women aged 20-64 relative to the total number of women in the same age group.

provided by ISTAT for the years 2018-2020, at the age and geographical area level. The years 2018-2020 were selected to align with the timing of the representative survey.

Italy's geographical disparities, which are among the most studied at the country level (see, among others, Bigoni et al. 2016, Putnam 1994, and Putnam 2000), offer a compelling backdrop for this analysis. These disparities are evident in labor market indicators, where northern regions typically outperform southern regions. In 2020, the overall employment rate for men in Italy was 71.8%, compared to 52.1% for women, highlighting a significant gender gap of almost 20 percentage points. Regionally, male employment rates ranged from 60.5% in the south to 78.9% in the north. The variation in female employment rates was even greater, ranging from 34.6% in the south to 62.6% in the north.

The lower employment rates for women in southern Italy reflect the much scarcer availability of childcare services, as noted by Del Boca (2002); Del Boca et al. (2004); and Del Boca and Saraceno (2005). This scarcity correlates positively with the documented relationship between mothers' labor supply and childcare provision; see De Henau et al. (2010).

Despite the limited sample size, the pronounced geographical heterogeneity within Italy provides valuable insights into the relevance of social norms analyzed in this study. We believe this can offer intriguing perspectives on the local influences shaping labor market dynamics for women.

With this objective in mind, we run the set of OLS regressions reported in Table **??**. Specifically, we regress employment rates from ISTAT administrative data and proxies of social norms calculated at the macro-area level using sample weight estimates. Following Fortin (2005), these social norm proxies are computed considering only male respondents to mitigate endogeneity issues. A description of the data used and their sources is provided in Table OA10 in the Online Appendix.

In all models, the dependent variable is the yearly employment rate, over three years, by age group and geographical area; this results in a total of 27 observations. Our list of controls includes the fraction of women holding a high school degree and the fraction holding a university degree at age and geographical area level for the years 2018–2020. In addition, as a proxy for regional spending on daycare services, we include the number of authorized places in public daycare per 100 children aged 0–2 years at the geographical area level for the years 2018–2020. Other controls for geographical macro-areas, age groups, and years are included in the analysis.

Model	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Female employment rate for Italian women aged 20–64 retrieved from ISTAT				
University degree	0.008*	0.007*	0.007	0.006*	0.007
	(0.0030)	(0.0027)	(0.0037)	(0.0027)	(0.0037)
High school degree	0.001	-0.001	-0.001	-0.001	-0.001
	(0.0021)	(0.0020)	(0.0027)	(0.0020)	(0.0028)
Childcare	0.007	0.008	0.008	0.008	0.008
	(0.0290)	(0.0217)	(0.0282)	(0.0214)	(0.0281)
Proxy of Social Norms					
Fraction SA/VA (HoB)		0.191**			
		(0.0500)			
Fraction SA/VA (FoB)			0.143		
			(0.0968)		
Mean rating (HoB)				0.179**	
				(0.0452)	
Mean Rating (FoB)					0.095
					(0.0630)
Geographical Area (Baseline: South and Islands)					
North	0.137	0.119	0.136	0.128	0.130
	(0.5017)	(0.3728)	(0.4863)	(0.3687)	(0.4846)
Centre	0.048	0.036	0.064	0.045	0.054
	(0.5946)	(0.4442)	(0.5765)	(0.4392)	(0.5744)
Age Groups (Baseline: 50–64)					
25–34	-0.145	-0.133	-0.093	-0.113	-0.091
	(0.0761)	(0.0700)	(0.0947)	(0.0698)	(0.0951)
35–49	0.006	0.021	0.034	0.038	0.044
	(0.0487)	(0.0438)	(0.0580)	(0.0441)	(0.0615)
Wave (Baseline: 2019)					
2020	-0.032*	-0.031**	-0.030*	-0.030**	-0.030*
	(0.0121)	(0.0093)	(0.0123)	(0.0095)	(0.0121)
2021	-0.037	-0.034	-0.031	-0.031	-0.031
	(0.0303)	(0.0247)	(0.0306)	(0.0251)	(0.0304)
Constant	0.107	0.070	0.096	0.171	0.172
	(0.4405)	(0.3299)	(0.4190)	(0.3279)	(0.4184)
Observations	27	27	27	27	27

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: **Association between elicited social norms and female employment rates**

Results from OLS regression. The dependent variable is the female employment rate retrieved from ISTAT for the years 2018–2020 at the age and geographical area level. "High school degree" and "University degree" refer to the share of women holding each degree. "Childcare" is the number of authorized public daycare slots per 100 children aged 0–2. Social norms are proxied via first- and higher-order beliefs, following Fortin (2005) and KW. All models include controls for macro-regions, age groups, and years. Robust standard errors in parentheses.

In models 2–5 of Table 12, we include proxies for social norms. As mentioned earlier, as a proxy for social norms, Fortin (2005) uses responses to statements from the WVS that elicit *male* respondents' personal opinions (first-order beliefs). Recall that we also gathered personal opinions in our representative survey. Specifically, for each of the three scenarios presented in the vignettes, respondents were explicitly asked to evaluate—according to their *personal opinion*—whether the behavior of the individual proposing the allocation was socially appropriate. Hence, to assess the external validity of social norms elicited as first-order beliefs (as in Fortin 2005) versus higher-order beliefs (following KW's methodology), we use two proxies based on first-order beliefs (see models (3) and (5)) and two proxies based on higher-order beliefs (see models (2) and (4)) in Table 12.

In models (2) and (3), social norms are defined as the fraction of male respondents who rate the equal contribution scenario in Vignette Part-Time as Somewhat Appropriate or Very Appropriate, based on higher-order and first-order beliefs, respectively. In models (4) and (5), social norms are instead defined as the mean appropriateness rating among male respondents in Vignette Part-Time, again using second-order and first-order beliefs, respectively.

Using either the fraction of respondents or the mean appropriateness rating yields similar qualitative results. However, only the social norm proxies based on higher-order beliefs show a significant association with the female employment rate (see models (2) and (4)). In contrast, proxies based on first-order beliefs do not reach statistical significance. This analysis indicates a positive and significant association between perceived approval (i.e., respondents' higher-order beliefs) for equal sharing of household responsibilities and female employment.
As expected, all models indicate that the share of women with a university degree is positively associated with female employment. In general, higher educational attainment translates into improved labor market opportunities. In addition, the year 2020 shows a negative association with the female employment rate, likely reflecting the impact of the COVID-19 pandemic.

Although not causal, these findings suggest that measures of social norms based on KW's methodology have explanatory power. Furthermore, they highlight the crucial link between gender equality in the household and women's participation in the labor market.

6 Conclusions

Using a representative survey of the Italian population (N=1,501), we elicit social norms as incentivized beliefs about others' beliefs through the KW method. Our sample is representative with respect to gender, age, residence area, and education, i.e., individual characteristics typically affecting perceptions of gender norms. As for respondents' age, representativeness holds across three age groups, 25–34, 35–59, and 50–64, that we use to compare gender norms across generations.
Our study includes two vignettes depicting hypothetical scenarios for a couple in which the female partner may work full-time or part-time, as well as a model in which partners contribute time to a family public good and experience disutility when deviating from a shared norm regarding socially approved divisions of domestic chores.
By embedding the KW incentivized methodology into a large-scale survey, we provide a cost-effective, rigorous, and behaviorally validated approach that integrates the strengths of experimental economics and empirical social research while preserving the theoretical foundation of social norms as collective expectations.

When partners in the vignette have similar labor market conditions, survey respondents evaluate equal contributions to household chores as socially appropriate. This

contrasts with both their self-reported behaviors in our representative survey and well-established time-use survey evidence, which consistently show that women devote significantly more time to household chores than their partners, even when both work full-time.

To understand this dissonance, we examine chore allocations that deviate from equality and identify two emerging biases. The first bias we document is a strong *framing effect*: a woman who proposes a self-beneficial allocation faces greater social stigma than a man making the same proposal. This suggests that social approval of the same chores allocation depends on the gender of the proposer. Additionally, women are perceived as more socially appropriate when taking primary responsibility for household chores (i.e., offering allocations that benefit their partner), whereas men exhibiting the same behavior do not face the same level of approval.

Beyond framing effects, we also document a clear *gender double standard*: a woman contributing less than her partner is perceived as less socially appropriate than a man in the same situation. However, this holds true only when she is the one initiating the allocation of chores. The two documented biases may reinforce traditional expectations about household responsibilities and help explain why traditional gender norms persist in time-use data. Interestingly, the younger generation is less affected by the double standard than older generations and appears entirely unaffected by the framing effect bias, suggesting a shift toward a more egalitarian norm when both partners work full-time.

When the female partner in the vignette works part-time, the likelihood of perceiving equal contributions to household chores as the social norm significantly decreases across elder age groups, further suggesting that younger generations are less influenced by traditional gender norms.

Finally, we provide evidence of a positive association between social norms measured using KW's methodology and female labor market participation as captured by Italian administrative data. This lends predictive power to our measure of gender norms and suggests that shared norms (beliefs about others' beliefs) can reflect views that influence—or are influenced by—societal patterns even better than personal values (first-order beliefs).

As a caveat, while our representative sample allows us to elicit gender equality norms across generations as they currently stand, it does not enable us to determine whether these differences represent permanent shifts. Thus, we cannot exclude the possibility that the youngest generation may adopt less progressive norms as they grow older and experience major life transitions, such as family formation. However, mapping these differences remains crucial, as norms influence individual behavior and societal expectations at each life stage.

To conclude, while our findings highlight the persistence of implicit biases in how fairness in household labor is perceived, more research is needed to understand how these biases evolve over time and what factors contribute to norm change. For example, among young adults, the reduced labor market participation of the female partner working part-time

may be viewed as a temporary and undesired outcome of labor market frictions rather than an acceptable justification for an unequal distribution of household work. Future research should explore whether these perceptions can lead to actual behavioral changes in the division of domestic labor.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGTP 5 to improve language and readability, with caution. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

References

Alesina, A., Miano, A., and Stantcheva, S. (2023). Immigration and redistribution. *The Review of Economic Studies*, 90(1):1–39.

Anxo, D., Mencarini, L., Pailhé, A., Solaz, A., Tanturri, M. L., and Flood, L. (2011). Gender differences in time use over the life course in france, italy, sweden, and the us. *Feminist economics*, 17(3):159–195.

Barigozzi, F., Di Timoteo, C., and Monfardini, C. (2023). The gender gaps in time-use within italian households during 2002–2014. *Italian Economic Journal*, 9(3):1263–1296.

Barigozzi, F. and Montinari, N. (2025). Social norms in survey experiments: Personal beliefs versus normative expectations. Technical report, Working Paper DSE N°1182, ISSN 2282-6483.

Barr, A., Lane, T., and Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, 164:153–164.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bertrand, M., Cortes, P., Olivetti, C., and Pan, J. (2021). Social norms, labour market opportunities, and the marriage gap between skilled and unskilled women. *The Review of Economic Studies*, 88(4):1936–1978.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.

Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., and Pancotto, F. (2016). Amoral familism, social capital, or trust? the behavioural foundations of the italian north–south divide. *The Economic Journal*, 126(594):1318–1341.

Boneva, T., Brás-Monteiro, A., Golin, M., and Rauh, C. (2024). Are men's preferences for couple equity misperceived? evidence from six countries. Technical report, IZA Discussion Paper N°17493.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.

Burks, S. V. and Krupka, E. L. (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science*, 58(1):203–217.

Bursztyn, L., Cappelen, A. W., Tungodden, B., Voena, A., and Yanagizawa-Drott, D. H. (2023). How are gender norms perceived? Technical report, National Bureau of Economic Research, Working Paper 31049, Revision date: February 2024.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029.

Bursztyn, L. and Yang, D. Y. (2022). Misperceptions about others. *Annual Review of Economics*, 14(1):425–452.

Cappadozzi, T. (2019). I tempi della vita quotidiana. lavoro, conciliazione, parità di genere e benessere soggettivo.

Carpenter, S. M. (2022). Framing effects. In *Encyclopedia of animal cognition and behavior*, pages 2793–2799. Springer.

Chang, D., Chen, R., and Krupka, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, 116:158–178.

Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

Cortés, P., Koşar, G., Pan, J., and Zafar, B. (2024). Should mothers work? how perceptions of the social norm affect individual attitudes toward work in the u.s. *The Review of Economics and Statistics*, pages 1–28.

Craig, L. and Mullan, K. (2011). How mothers and fathers share childcare: A cross-national time-use comparison. *American sociological review*, 76(6):834–861.

d'Adda, G., Drouvelis, M., and Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62:1–7.

De Henau, J., Meulders, D., and O'Dorchai, S. (2010). Maybe baby: Comparing partnered women's employment and child policies in the eu-15. *Feminist economics*, 16(1):43–77.

Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in italy. *Journal of population economics*, 15:549–573.

Del Boca, D., Pasqua, S., and Pronzato, C. (2004). Employment and fertility decisions in italy france and the uk. *Centre for Household, Income, Labour and Demographic Economics Working Papers*, pages 8–04.

Del Boca, D., Pasqua, S., and Suardi, S. (2015). Childcare, mothers' work and children's schooling outcomes. an analysis of italian data. *Families and Societies Working Paper Series*.

Del Boca, D. and Saraceno, C. (2005). Le donne in italia tra famiglia e lavoro. *Economia & lavoro*, 39(1):125–0.

Erkut, H., Nosenzo, D., and Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, 130:28–31.

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691.

Fernández, R. and Fogli, A. (2005). Culture: An empirical investigation of beliefs, work, and fertility.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across oecd countries. *oxford review of Economic Policy*, 21(3):416–438.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42.

Fromell, H., Nosenzo, D., Owens, T., and Tufano, F. (2021). One size does not fit all: Plurality of social norms and saving behavior in kenya. *Journal of Economic Behavior & Organization*, 192:73–91.

Gächter, S., Gerhards, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97:72–86.

Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3):548–573.

Gimenez-Nadal, J. I. and Molina, J. A. (2020). The gender gap in time allocation in europe.

Gosling, S. D., Rentfrow, P. J., and Swann Jr, W. B. (2003). Ten-item personality inventory. *Journal of Research in Personality*.

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimuller, J. (2019). Child penalties across countries: Evidence and explanations. In *AEA Papers and Proceedings*, volume 109, pages 122–26.

König-Kersting, C. (2021). On the robustness of social norm elicitation. Technical report, Working Papers in Economics and Statistics.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Nosenzo, D. and Görges, L. (2020). Measuring social norms in economics: Why it is important and how it is done. *Analyse & Kritik*, 42(2):285–312.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of economic perspectives*, 14(3):137–158.

Putnam, R. D. (1994). Making democracy work: Civic traditions in modern italy.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and schuster.

Ridgeway, C. L. and Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & society*, 18(4):510–531.

Schneeberger, A. and Krupka, E. L. (2021). Determinants of norm compliance: Moral similarity and group identification. *Available at SSRN 3969227*.

Settele, S. (2022). How do beliefs about the gender wage gap affect the demand for public policy? *American Economic Journal: Economic Policy*, 14(2):475–508.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15(1):205–234.

Thébaud, S., Kornrich, S., and Ruppanner, L. (2021). Good housekeeping, great expectations: Gender and housework norms. *Sociological Methods & Research*, 50(3):1186–1214.

Trappe, H., Pollmann-Schult, M., and Schmitt, C. (2015). The rise and decline of the male breadwinner model: Institutional underpinnings and future expectations. *European Sociological Review*, 31(2):230–242.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *science*, 211(4481):453–458.

Veselỳ, Š. (2015). Elicitation of normative and fairness judgments: Do incentives matter? *Judgment and Decision making*, 10(2):191–197.

von Gleichen, R. D. and Seeleib-Kaiser, M. (2018). Family policies and the weakening of the male-breadwinner model. In *Handbook on gender and social policy*, pages 153–178. Edward Elgar Publishing.

Appendix

Activity	Men	Women	p-value
Shopping	18.34	71.27	0.000
	(.0232)	(.0281)	
Wash and Iron	45.65	84.25	0.000
	(.0299)	(.0226)	
Cooking	36.67	78.64	0.000
	(.0289)	(.0254)	
Cleaning	39.76	71.12	0.000
	(.0294)	(.0281)	

Table A1: **Subsample of married or cohabiting and working individuals** (N = 705 out of 1,501). Percentage reporting that *the female partner executes more often/always household activities in the couple* (Std. Error in parenthesis).

Note: For each activity, we report the share of respondents indicating that the woman is primarily responsible (i.e. executing the task always or the most part of the times). Specifically, for female respondents, we report the percentage who say they perform the task exclusively or more often than their partner. For male respondents, we report the percentage who say their partner takes care of the task more often or always. Thus, while the indicator is consistent in concept —identifying when women bear greater responsibility— its definition varies depending on whether it is self-reported or reported by the partner. In each panel, the reported percentages are calculated using the Stata command .svy, which adjusts for sample weights. As a result, these percentages may differ slightly from those obtained using unweighted group sizes. All tests were adjusted for multiple comparisons using the Benjamini-Hochberg False Discovery Rate correction (Benjamini Hochberg, 1995).

Activity	Men	Women	p-value
Age Group: 25-34			
Shopping	9.91	51.92	0.000
	(.0414)	(.0780)	
Wash and Iron	22.65	83.25	0.000
	(.0580)	(.0583)	
Cooking	26.05	72.62	0.000
	(.0609)	(.0696)	
Cleaning	30.01	58.70	0.000
	(.0636)	(.0769)	
Age Group: 35-49			
Shopping	20.79	77.74	0.000
	(.0298)	(.0304)	
Wash and Iron	46.52	84.57	0.000
	(.0365)	(.0264)	
Cooking	38.47	77.22	0.000
	(.0357)	(.0307)	
Cleaning	39.53	72.98	0.000
	(.0358)	(.0325)	
Age Group: 50-64			
Shopping	18.23	74.08	0.000
	(.0623)	(.0800)	
Wash and Iron	61.00	84.41	0.000
	(.0791)	(.0662)	
Cooking	39.99	89.24	0.000
	(.0795)	(.0566)	
Cleaning	47.86	79.43	0.000
	(.0810)	(.0738)	

Table A2: Percentage reporting that the women execute more often/always household activities by age group (Std. Error in parenthesis)

Note: In each panel, the reported percentages are calculated using the Stata command .svy, which adjusts for sample weights. As a result, these percentages may differ slightly from those obtained using unweighted group sizes. The analysis is restricted to respondents who are in a couple (married or cohabiting and working, N = 705/1,501). For each activity, we report the share of respondents indicating that the woman is primarily responsible (i.e. executing the task always or the most part of the times). Specifically, for female respondents, we report the percentage who say they perform the task exclusively or more often than their partner. For male respondents, we report the percentage who say their partner takes care of the task more often or always. Thus, while the indicator is consistent in concept —identifying when women bear greater responsibility— its definition varies depending on whether it is self-reported or reported by the partner. All tests were adjusted for multiple comparisons using the Benjamini-Hochberg False Discovery Rate correction (Benjamini Hochberg, 1995).

Table A3: Summary Statistics

Variable	Frequency		Description
Female	58.43		Female respondent
Male	41.57		Male respondent
North	47.90		Geographical area of residence
Centre	18.92		Geographical area of residence
South and Islands	33.18		Geographical area of residence
25-34	19.85		Age group
34-49	52.43		Age group
50-64	27.71		Age group
Woman	54.10		Proposer's gender
Man	45.90		Proposer's gender
Children	58.63		Respondent has children
Married or Cohabitant	72.15		Respondent is married/ cohabiting
University Degree	35.38		Respondent has a tertiary degree
Employed	63.82		Respondent is working
Free Time	16.85		Most important life dimension
Community Involvement	3.93		Most important life dimension
Work	22.78		Most important life dimension
Family	70.55		Most important life dimension
Centre Right[25]	24.38		Political orientation
Cognitive Reflection Test	13.26		Two out of three correct answers to the CRT
Risk Attitude	58.36		Above the median[26] attitude towards risk
Trust Attitude	16.66		Respondent trusts most people
Strongly Agree	6.06		To claim 5
Agree	29.91		To claim 5
Disagree	28.98		To claim 5
Strongly Disagree	35.04		To claim 5
Less than 5,000	4.55		Inhabitants of the city of residence
Between 5,000 and 10,000	7.99		Inhabitants of the city of residence
Between 10,000 and 50,000	53.9		Inhabitants of the city of residence
Variable	**Mean**	**sd**	**Description**
Big 5 Personality Traits			
Agreeableness	5.31	1.09	Good-natured, cooperative, trustful
Conscientiousness	5.52	1.12	Orderly, responsible, dependable
Emotional Stability	4.54	1.24	Calm, non neurotic, non easily upset
Openness to Experience	4.28	1.05	Intellectual, imaginative, independent-minded
Extraversion	4.00	1.37	Talkative, assertive, energetic
Gini Index	.21	0.02	Provice of residence's gini index[27]
Male/Female Employment Ratio	1.52	0.29	Province of residence's ratio male to female employment ratio (employed wrt the resident population of 15 years or more).[28]

We report frequencies for categorical variables; mean and standard deviations for continuous variables included in the analysis. A description is presented for each variable together with the source for those that were not surveyed.

[25]Centre-right comprehends: Lega, Forza Italia, and Fratelli d'Italia (respective shares: 56.01%, 19.13%, 24.86%).

[26]Risk Attitude has a median of 6 in a scale where 0 stands for "absolutely not willing to take risks"and 10 stands for "absolutely willing to take risks".

[27]Source: urbanindex.it; Atlante PRIN Postmetropoli, elaborazioni su dati MEF - Ministero dell'Economia e della Finanza.

[28]Source: ISTAT 8milaCensus and own calculations.

Table A4: Randomization check

Variable	Man Proposing	Woman Proposing	Mean Difference	p-value
Female	55.42	61.97	.066	0.3472
Male	38.03	44.58	-.066	0.1736
Geographical Area				
North	48.33	47.54	0.008	0.9218
Center	19.45	18.47	0.010	0.9326
South and Isles	32.22	33.99	-.018	0.9954
Age Group				
25-34	20.90	18.97	0.019	0.8488
35-49	50.94	53.69	-.028	0.9787
50-64	28.16	27.34	0.008	0.9127
Children	59.36	58.00	.014	0.9636
Married or Cohabitant	72.86	71.55	.013	1.0000
University Degree	36.44	34.48	.019	0.9795
Working	60.81	66.38	-.056	0.2152
Important dimensions in life				
Free Time	16.40	17.24	-.008	0.9418
Community Involvement	4.06	3.82	.002	0.9146
Work	22.35	23.15	-.008	0.9689
Family	69.81	71.18	-.014	1.0000
Political orientation				
Centre Right	26.27	22.78	.035	0.5688
Personality Traits				
Cognitive Reflection Test	13.79	18.35	-.046	0.1924
Risk Attitude	57.62	58.99	-.014	1.0000
Trust Attitude	15.53	17.61	-.021	1.0000
Big Five Personality Traits				
Agreeableness	5.30	5.32	-.020	0.939
Conscientiousness	5.46	5.57	-.118	0.2822
Emotional Stability	452	4.56	-.042	1.0000
Openness to Experience	4.29	4.28	.008	0.9157
Extroversion	4.04	3.96	.084	0.9959
Claim: A woman should be ready to reduce the time devoted to her job for family reasons				
Strongly Agree	6.10	6.03	.001	0.9605
Agree	29.61	30.17	-.006	0.8628
Disagree	28.59	29.31	-.007	0.8911
Strongly Disagree	35.70	34.48	.012	0.9605
Municipality size: Inhabitants				
Less than 5,000	3.92	5.06	-.011	1.0000
Between 5,000 and 10,000	8.70	7.38	.013	0.9862
Between 10,000 and 50,000	53.56	54.19	-.006	0.8851
ISTAT data at the municipality level				
Gini Index	0.21	0.21	-.001	0.9126
Male/Female Employment Ratio	1.51	1.54	-.027	0.5627

We report frequencies for categorical variables; mean and standard deviations for continuous variables. Benjamini-Hochberg adjusted p-values are presented, the p-values refer to a test of equality of means between woman-proposing and man-proposing samples.

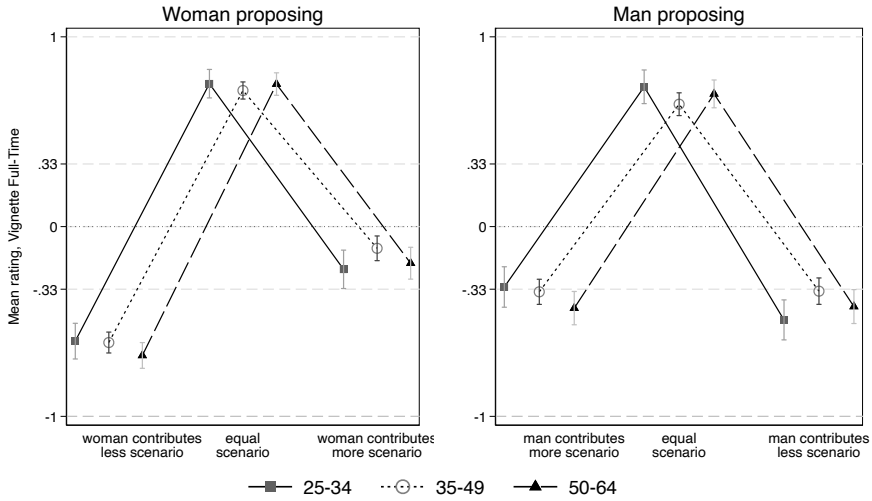

Figure A1: Norm function for Vignette B: "Imagine Antonio and Francesca: they are married or cohabiting. They both work the same number of hours, earn roughly the same amount of money, and have the same career trajectories. They have no children and no one to help them with the housework." The plot on the left represents the norm function for the respondents exposed to the "Woman proposing" treatment, and on the right is the norm function for the respondents exposed to the "Man proposing" treatment. 95% Confidence intervals are displayed.

	Woman Contributes Less			Equal Contribution			Man Contributes Less		
	25-34	35-49	50-64	25-34	35-49	50-64	25-34	35-49	50-64
Very Socially Inappropriate	45.49	47.06	**53.46**	1.3	2.29	1.63	37.1	28.76	33.73
Somewhat Socially Inappropriate	33.72	34.12	30.44	5.46	6.44	6.51	34.78	33.51	34.39
Somewhat Socially Appropriate	16.38	14.76	13.17	23.28	27.33	22.97	21.97	28.7	25.07
Very Socially Appropriate	4.41	4.05	2.93	**69.96**	**63.95**	**68.88**	6.14	9.03	6.82
Mean Rating	-.4679	-.494	-.5624	.7453	.6856	.7268	-.3519	-.2132	-.2999

Panel a) Mean Differences, gender double standard within (between)

	25-34 (vs 35-49)	35-49 (vs 50-64)	50-64 (vs 25-34)
p-value	.0097 (.0069)	.0000 (.7131)	.0000 (.0213)

Panel b) Mean Differences, within scenario between generations

25-34 vs 35-49	.0261	(.6140)	.0597	(.1558)	-.1387	(.0218)
25-34 vs 50-64	.0945	(.1310)	.0185	(.6248)	-.052	(.3832)
35-49 vs 50-64	.0684	(.1323)	-.0412	(.2595)	.0867	(.1444)

Table A5: Family norm by age groups, Vignette B

Vignette B: "Imagine Antonio and Francesca: they are married or cohabiting. They both work the same number of hours, earn roughly the same amount of money, and have the same career trajectories. They have no children and no one to help them with the housework.". The elicited social norm is presented inside a box, and strong norms (i.e., norms shared by the majority of the sample) are presented in boldface.
In Panel a), Benjamini-Hochberg adjusted p-values referring to a test of equality of means in the woman and man contributes less scenario within age groups and between age groups are shown in parenthesis (results are replicated with Wilcoxon rank-sum test.). In Panel b), Benjamini-Hochberg adjusted p-values in parenthesis, the p-values refer to a test of equality between age groups in each scenario.

Table A6: Family norm by age groups, Vignette A

	Woman Contributes Less			Equal Contribution			Man Contributes Less		
	25-34	35-49	50-64	25-34	35-49	50-64	25-34	35-49	50-64
Very Socially Inappropriate	**54.61**	**56.71**	**59.42**	6.13	10.72	13.68	9.96	9.09	10.98
Somewhat Socially Inappropriate	26.21	25.19	25.78	30.61	31.71	37.25	15.56	17.11	18.77
Somewhat Socially Appropriate	14.5	13.96	12.22	39.54	36.99	29.54	42.71	41.88	39.25
Very Socially Appropriate	4.69	4.14	2.58	23.72	20.59	19.53	31.77	31.92	31
Mean Rating	-.5379	-.5628	-.6131	.2054	.1161	.033	.3077	.3101	.2678

Mean Differences

25-34 vs 35-49	.0249	(.6421)	.0893	(.1911)	-.0024	(.9588)
25-34 vs 50-64	.0752	(.241)	.1724	(.0046)	.0399	(.5721)
35-49 vs 50-64	.0503	(.2887)	.0831	(.1292)	.0423	(.2887)

Vignette A: "Imagine Giulio and Silvia: they are married or cohabiting. Giulio works twice as many hours as Silvia and earns about twice as much. They have no children and no one to help them with the housework.". The elicited social norm is presented inside a box, strong norms (i.e. norms shared by the majority of the sample) are presented in boldface. Benjamini-Hochberg adjusted p-values in parenthesis, the p-values refer to a test of equality between age groups in each scenario.

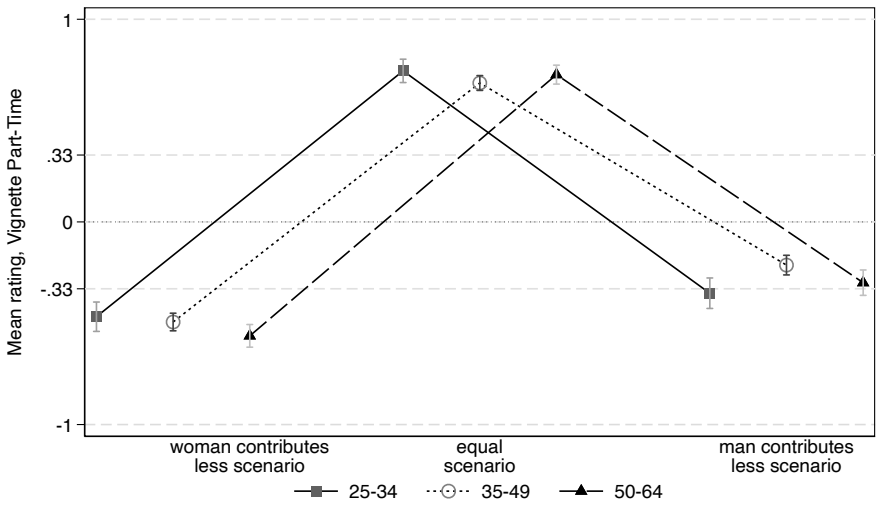

Figure A2: Norm function for Vignette B: "Imagine Antonio and Francesca: they are married or cohabiting. They both work the same number of hours, earn roughly the same amount of money, and have the same career trajectories. They have no children and no one to help them with the housework." The solid line represents the norm function for the younger generation, the dashed line represents the norm function for the middle-aged generation, and the dotted line represents the norm function for the older generation. 95% Confidence intervals are displayed.

	(1)	(2)	(3)	(4)
Dependent Variable	1 if identifies a norm punishing the "Woman contributes less" more than the "Man contributes less" scenario, 0 otherwise			
Woman Proposing				
Female	-0.048	-0.050	-0.055	-0.072
	(0.0393)	(0.0399)	(0.0399)	(0.0425)
North	-0.076	-0.086	-0.086	-0.080
	(0.0446)	(0.0469)	(0.0468)	(0.0478)
Centre	0.013	-0.010	-0.008	-0.003
	(0.0569)	(0.0596)	(0.0593)	(0.0598)
25-34	-0.044	-0.045	-0.009	-0.014
	(0.0572)	(0.0576)	(0.0617)	(0.0620)
35-49	-0.019	-0.026	-0.021	-0.020
	(0.0438)	(0.0447)	(0.0447)	(0.0448)
Moved North		0.021	0.016	0.011
		(0.0615)	(0.0614)	(0.0616)
Moved South		-0.053	-0.067	-0.069
		(0.1034)	(0.1015)	(0.1005)
Having Children			0.032	0.031
			(0.0478)	(0.0478)
Married or Cohabitant			0.089	0.094
			(0.0492)	(0.0491)
Controls				
Education and Job				✓
Observations	812	786	786	786

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: Vignette 2, Woman Proposing Survey Logit AME

Average marginal effects for the probability of rating the woman contributing less scenario less appropriate than the man contributing less scenario in the woman proposing subsample. **In columns (2)-(4) we loose data on 26 observations as we do not have reliable information on the geographical area at birth, or as respondents were born abroad.**

	(1)	(2)	(3)	(4)
Dependent Variable	1 if identifies a norm punishing the "Woman contributes less" more than the "Man contributes less" scenario, 0 otherwise			
Man Proposing				
Female	0.059	0.078*	0.076*	0.066
	(0.0365)	(0.0366)	(0.0370)	(0.0391)
North	-0.016	-0.012	-0.010	-0.006
	(0.0426)	(0.0441)	(0.0440)	(0.0436)
Centre	-0.009	-0.021	-0.020	-0.020
	(0.0523)	(0.0543)	(0.0546)	(0.0542)
25-34	-0.018	-0.029	-0.017	-0.023
	(0.0508)	(0.0503)	(0.0560)	(0.0561)
35-49	-0.016	-0.025	-0.018	-0.020
	(0.0409)	(0.0415)	(0.0424)	(0.0429)
Moved North		-0.017	-0.012	-0.015
		(0.0533)	(0.0544)	(0.0539)
Moved South		0.176	0.174	0.176
		(0.1102)	(0.1107)	(0.1114)
Having Children			0.038	0.043
			(0.0458)	(0.0463)
Married or Cohabitant			-0.026	-0.028
			(0.0488)	(0.0490)
Controls				
Education and Job				✓
Observations	689	657	657	657

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A8: Vignette 2, Man Proposing Survey Logit AME

Average marginal effects for the probability of rating the woman contributing less scenario less appropriate than the man contributing less scenario in the man proposing subsample. **In columns (2)-(4) we loose data on 32 observations as we do not have reliable information on the geographical area at birth, or as respondents were born abroad.**

Online Appendix

Table OA1: Representativeness

	Italian Population			Survey Sample		
Age Range	**25-34**	**35-49**	**50-64**	**25-34**	**35-49**	**50-64**
			North-West			
Males	2.58	5.33	5.36	2.56	5.30	5.33
Females	2.48	5.25	5.52	2.48	5.26	5.52
Overall	5.06	10.58	10.87	5.05	10.55	10.85
			North-East			
Males	1.86	3.88	3.92	1.84	3.84	3.89
Females	1.81	3.85	4.03	1.81	3.83	4.01
Overall	3.67	7.73	7.95	3.64	7.67	7.91
			Centre			
Males	1.94	3.94	3.91	1.94	3.94	3.90
Females	1.88	4.05	4.19	1.90	4.07	4.21
Overall	3.82	8.00	8.10	3.84	8.01	8.11
			South and Islands			
Males	3.81	6.55	6.53	3.83	6.56	6.54
Females	3.67	6.66	6.99	3.71	6.70	7.03
Overall	7.48	13.22	13.52	7.54	13.25	13.57

Data extraction: April 21^{st} 2023 from I.Stat. Reference period: 2019

Table OA2: Survey text

Participants were shown the following text (here translated from Italian):
"When answering the next 5 questions, you can win an Amazon voucher if you guess the answer chosen by most people *similar to you* who are responding to this survey. By similar to you, we mean: of your same gender, in your age group (i.e., AGE GROUP), and residing in your same geographical area (i.e., AREA)." "When all participants have completed the questionnaire, we will conduct two drawings: 1) We will randomly select 1 out of the next 5 questions. 2) We will randomly select 150 participants from those who have completed the survey (out of 1500 people). "Among the 150 selected, those who correctly guessed the answer given by the majority of other participants similar to them on the selected question will receive 3 euros for each correct answer. The amount earned by each of the selected participants will be sent by Scenari Srl."

At the beginning of the elicitation part, participants were presented the following text (here translated from Italian):
"In the next 4 questions, you will read descriptions of situations where a couple has to decide how to organize the management of household tasks and childcare. For each situation, you will be given a brief description of the partners' jobs and the possible solutions they have adopted. You will be asked to evaluate different organizational choices made by the partners of a couple, indicating for each one whether most people similar to you would consider them "socially appropriate" or "socially inappropriate". By "socially appropriate" organizational choices, we mean family decisions that most people agree are the "correct" or "right" thing to do. Another way to think about what we mean is that if someone organizes their family life in a socially appropriate way, then no one else can judge that person negatively for their choices."

Table OA3: Gender double standard, Vignette B

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	1 if identifies a norm punishing the "Woman contributes less" more than the "Man contributes less" scenario, 0 otherwise					
Independent Variables						
Female	-0.052	-0.035	-0.041	-0.042	-0.034	-0.046
	(0.0275)	(0.0280)	(0.0281)	(0.0304)	(0.0300)	(0.0322)
Age Groups (Baseline: 50-64)						
25-34	-0.078*	-0.078*	-0.049	-0.040	-0.037	-0.009
	(0.0376)	(0.0381)	(0.0416)	(0.0418)	(0.0404)	(0.0445)
35-49	0.024	0.021	0.026	0.033	0.017	0.055
	(0.0311)	(0.0317)	(0.0320)	(0.0318)	(0.0307)	(0.0333)
Geographical Areas (Baseline: South and Islands)						
North	-0.021	-0.018	-0.016	-0.009	-0.018	-0.021
	(0.0310)	(0.0331)	(0.0329)	(0.0332)	(0.0331)	(0.0488)
Centre	-0.020	-0.038	-0.033	-0.028	-0.036	-0.034
	(0.0399)	(0.0418)	(0.0417)	(0.0419)	(0.0403)	(0.0520)
Relocated to a different Geographical Area (Baseline: Did not move)						
Moved North		0.038	0.038	0.042	0.054	0.039
		(0.0423)	(0.0429)	(0.0427)	(0.0419)	(0.0447)
Moved South		0.072	0.064	0.061	0.068	0.016
		(0.0713)	(0.0704)	(0.0704)	(0.0678)	(0.0734)
Civil Status (Baseline: Single, Widower, Separated-Divorced)						
Married or Cohabitant			0.049	0.052	0.036	0.047
			(0.0344)	(0.0343)	(0.0341)	(0.0361)
Having Children			0.043	0.039	0.027	0.048
			(0.0336)	(0.0336)	(0.0331)	(0.0365)
Framing: Woman Proposing	0.276***	0.277***	0.277***	0.277***	0.269***	0.257***
	(0.0273)	(0.0277)	(0.0276)	(0.0275)	(0.0267)	(0.0286)
University Degree				-0.077**	-0.063*	-0.071*
				(0.0275)	(0.0279)	(0.0293)
Employed				-0.016	0.008	-0.008
				(0.0323)	(0.0318)	(0.0355)
Important spheres of life						
Free time					-0.007	0.014
					(0.0435)	(0.0460)
Community Involvement					-0.002	0.024
					(0.0690)	(0.0738)
Work					0.015	-0.009
					(0.0390)	(0.0415)
Family					0.069	0.045
					(0.0382)	(0.0413)
Centre right					0.051	0.047
					(0.0325)	(0.0349)
TIPI						
Agreeableness					0.000	0.008
					(0.0142)	(0.0157)
Conscientiousness					0.011	0.009
					(0.0135)	(0.0145)
Emotional stability					-0.012	-0.017
					(0.0120)	(0.0126)
Openness					0.003	-0.000
					(0.0144)	(0.0155)
Extraversion					-0.019	-0.017
					(0.0107)	(0.0115)
Cognitive Reflection Test						
2 correct answers					0.087*	0.101**
					(0.0364)	(0.0387)
Risk attitude above median					0.027	0.028
					(0.0289)	(0.0308)
Trust time most of the time					-0.010	-0.032
					(0.0358)	(0.0383)

Claim[29] **(Baseline: Strongly Agree)**						
Claim 5 A					-0.018	-0.024
					(0.0599)	(0.0666)
Claim 5 D					-0.118	-0.114
					(0.0607)	(0.0668)
Claim 5 SD					-	-
					0.237***	0.241***
					(0.0598)	(0.0664)
Municipality inhabitants (Baseline: More than 50,000)						
Less than 5,000						0.123
						(0.0914)
Between 5,000 and 10,000						-0.052
						(0.0693)
Between 10,000 and 50,000						0.026
						(0.0443)
Gini index						-0.225
						(0.9688)
Male to female employment ratio						-0.067
						(0.0678)
Observations	1501	1443	1443	1443	1443	1243

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Average marginal effects for the probability of rating the woman contributing less scenario less appropriate than the man contributing less scenario. In columns (2)-(4) we loose data on 58 observations as we do not have reliable information on the geographical area at birth, or as respondents were born abroad. In column (6) we loose additional 200 individuals since we are not able to match all municipalities in our dataset.

[29]"A woman should be ready to reduce the time devoted to her job for family reasons"

Table OA4: Gender double standard, models with interactions

	(1)	(2)	(3)
Dependent Variable	1 if identifies a norm punishing the "Woman contributes less" more than the "Man contributes less" scenario, 0 otherwise		
Panel a) AME for a change in age groups (baseline: 50-64)			
25–34			
Male	-0.099		
	(0.0565)		
Female	-0.058		
	(0.0497)		
North		-0.013	
		(0.0541)	
Centre		-0.050	
		(0.0932)	
South and Islands		-0.179**	
		(0.0623)	
North × Male			0.008
			(0.0815)
North × Female			-0.039
			(0.0710)
Centre × Male			0.008
			(0.1308)
Centre × Female			-0.107
			(0.1293)
South and Islands × Male			-0.309**
			(0.0962)
South and Islands × Female			-0.057
			(0.0791)
35–49			
Male	-0.001		
	(0.0486)		
Female	0.048		
	(0.0392)		
North		0.026	
		(0.0435)	
Centre		0.083	
		(0.0702)	
South and Islands		-0.013	
		(0.0565)	
North × Male			-0.055
			(0.0673)
North × Female			0.106
			(0.0550)
Centre × Male			0.237*
			(0.1004)
Centre × Female			-0.072
			(0.0958)
South and Islands × Male			-0.072
			(0.0922)
South and Islands × Female			0.041
			(0.0667)
Panel a) AME for a change in proposer's gender (baseline: Man proposing)			
Male			0.278***
			(0.0423)
Female			0.274***
			(0.0345)
Observations	1501	1501	1501

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Average marginal effects for the change in the probability of rating the woman contributing less scenario less appropriate than the man contributing less scenario with respect to age groups and proposer's gender.

Table OA5: Decline of the bread-winner model, Vignette A

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	1 if identifies Very or Somewhat Appropriate as norm in the equal share scenario, 0 otherwise					
Independent Variables						
Female	-0.019	-0.018	-0.015	-0.025	-0.019	-0.037
	(0.0291)	(0.0296)	(0.0297)	(0.0318)	(0.0319)	(0.0340)
Age Groups (Baseline: 50-64)						
25—34	0.143***	0.145***	0.131**	0.126**	0.119**	0.146**
	(0.0404)	(0.0409)	(0.0436)	(0.0439)	(0.0426)	(0.0456)
35–49	0.085**	0.092**	0.088**	0.087*	0.096**	0.104**
	(0.0327)	(0.0332)	(0.0337)	(0.0338)	(0.0331)	(0.0356)
Geographical Area of Residence (Baseline: South and Islands)						
North	0.055	0.038	0.036	0.040	0.028	0.051
	(0.0329)	(0.0350)	(0.0349)	(0.0356)	(0.0366)	(0.0543)
Centre	0.009	-0.016	-0.018	-0.016	-0.020	0.008
	(0.0424)	(0.0436)	(0.0436)	(0.0440)	(0.0429)	(0.0558)
Relocation to a different Geographical Area (Baseline: did not move)						
Moved North		0.044	0.040	0.038	0.038	0.064
		(0.0452)	(0.0453)	(0.0451)	(0.0435)	(0.0458)
Moved South		-0.063	-0.059	-0.059	-0.085	-0.056
		(0.0748)	(0.0756)	(0.0755)	(0.0777)	(0.0835)
Framing: Woman Proposing	0.009	0.003	0.003	0.005	0.013	-0.022
	(0.0291)	(0.0296)	(0.0296)	(0.0297)	(0.0290)	(0.0312)
Civil Status (Baseline: Single, Widower, Separated-Divorced)						
Married or Cohabitant			-0.002	-0.001	0.005	0.003
			(0.0368)	(0.0367)	(0.0362)	(0.0382)
Having Children			-0.031	-0.028	-0.011	-0.033
			(0.0353)	(0.0352)	(0.0349)	(0.0379)
University Degree				0.031	0.018	0.008
				(0.0305)	(0.0309)	(0.0327)
Employed				-0.022	-0.042	-0.038
				(0.0335)	(0.0334)	(0.0364)
Free time					-0.012	0.001
					(0.0439)	(0.0460)
Community Involvement					0.069	0.077
					(0.0724)	(0.0740)
Work					0.079*	0.100*
					(0.0397)	(0.0416)
Family					0.008	0.035
					(0.0406)	(0.0426)
Centre right					-0.012	-0.016
					(0.0346)	(0.0374)
TIPI						
Agreeableness					-0.005	-0.013
					(0.0154)	(0.0168)
Conscientiousness					-0.039**	-0.043**
					(0.0144)	(0.0156)
Emotional Stability					0.013	0.022
					(0.0129)	(0.0137)
Openness					0.005	0.005
					(0.0154)	(0.0172)
Extraversion					0.000	0.000
					(0.0114)	(0.0125)
Cognitive Reflection Test						
2 correct answers					-0.040	-0.020
					(0.0403)	(0.0429)
Risk attitude above median					0.014	0.003
					(0.0311)	(0.0328)
Trust most of the time					0.024	0.038
					(0.0388)	(0.0407)

Claim[30] **(Baseline: Strongly Agree)**						
Agree					-0.009	0.075
					(0.0684)	(0.0704)
Disagree					0.108	0.187**
					(0.0694)	(0.0710)
Strongly Disagree					0.195**	0.267***
					(0.0692)	(0.0706)
Municipality inhabitants (Baseline: More than 50,000)						
Less than 5,000						0.078
						(0.0898)
Between 5,000 and 10,000						0.123
						(0.0688)
Between 10,000 and 50,000						0.021
						(0.0493)
Gini index						0.630
						(1.0323)
Male to female employment ratio						0.092
						(0.0771)
Observations	1501	1443	1443	1443	1443	1243

*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Average marginal effects for the probability of rating the equality scenario Very Appropriate or Somewhat Appropriate. In columns (2)-(5) we loose 58 observations as we do not have reliable information on the geographical area of birth, or as respondents were born abroad. In column (6) we loose additional 200 individuals since we are not able to match all municipalities in our dataset.

[30]"A woman should be ready to reduce the time devoted to her job for family reasons"

Table OA6: Decline of the bread-winner model, models with interactions

Model	(1)	(2)	(3)
Dependent Variable	1 if identifies Very or Somewhat Appropriate as norm in the equal share scenario, 0 otw		
Panel a) AME for a change in age groups (baseline: 50–64)			
25–34			
Male	0.127*		
	(0.0610)		
Female	0.159**		
	(0.0528)		
North		0.165**	
		(0.0568)	
Centre		0.235*	
		(0.1002)	
South and Islands		0.063	
		(0.0686)	
North × Male			0.136
			(0.0842)
North × Female			0.198**
			(0.0755)
Centre × Male			0.075
			(0.1535)
Centre × Female			0.387**
			(0.1206)
South and Islands × Male			0.134
			(0.1038)
South and Islands × Female			-0.011
			(0.0893)
34–49			
Male	0.064		
	(0.0507)		
Female	0.105*		
	(0.0415)		
North		0.150***	
		(0.0454)	
Centre		0.088	
		(0.0738)	
South and Islands		-0.007	
		(0.0597)	
North × Male			0.197**
			(0.0689)
North × Female			0.104
			(0.0589)
Centre × Male			-0.043
			(0.1116)
Centre × Female			0.210*
			(0.0873)
South and Islands × Male			-0.059
			(0.0933)
South and Islands × Female			0.044
			(0.0751)
Panel b) AME for a change in proposer's gender (baseline: Man proposing)			
Male			0.026
			(0.0453)
Female			-0.008
			(0.0373)
Observations	1501	1501	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Average marginal effects for the change in the probability of rating the equality scenario Very Appropriate or Somewhat Appropriate with respect to age groups and proposer's gender.

Table OA7: Misperception in Vignette A

	(1)	(2)	(3)	(4)
Dependent Variable	1 if correctly identifies the norm in the "Equal contribution" scenario			
25–34	-0.057			
	(0.0405)			
35–49	-0.014			
	(0.0318)			
			25–34	
North	-0.016		-0.020	
	(0.0321)		(0.0571)	
Centre	-0.032		-0.060	
	(0.0414)		(0.1028)	
Female	0.004	-0.051		
	(0.0285)	(0.0536)		
Woman Proposing	-0.028			
	(0.0286)			
Male		-0.063		
		(0.0607)		
South and Islands			-0.098	
			(0.0677)	
North × Male				0.036
				(0.0821)
North × Female				-0.076
				(0.0787)
Centre × Male				-0.180
				(0.1508)
Centre × Female				0.054
				(0.1337)
South and Islands × Male				-0.117
				(0.1041)
South and Islands × Female				-0.077
				(0.0864)
			35–49	
Male		-0.035		
		(0.0491)		
Female		0.007		
		(0.0405)		
North			-0.039	
			(0.0450)	
Centre			0.010	
			(0.0710)	
South and Islands			0.007	
			(0.0573)	
North × Male				-0.055
				(0.0696)
North × Female				-0.023
				(0.0570)
Centre × Male				-0.080
				(0.1014)
Centre × Female				0.095
				(0.0926)
South and Islands × Male				0.021
				(0.0903)
South and Islands × Female				-0.006
				(0.0714)
Observations	1501	1501	1501	1501

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Column (1) presents the average marginal effects, columns (2)-(4) presents the average marginal effects for a change in age group (baseline: 50–64).

Table OA8: Vignette A, Self

	Woman Contributes Less			Equal Contribution			Man Contributes Less		
	25-34	35-49	50-64	25-34	35-49	50-64	25-34	35-49	50-64
Very Socially Inappropriate	48.35	49.71	55.24	5.31	10.67	10.35	11.44	10.11	12.76
Somewhat Socially Inappropriate	29.58	31.12	29.73	37.79	31.24	36.24	16.2	18.63	19.42
Somewhat Socially Appropriate	18.21	15.06	11.27	34.58	36.44	30.66	48.11	46.36	39.45
Very Socially Appropriate	3.87	4.1	3.75	22.33	21.65	22.75	24.25	24.9	28.37
Mean Rating	-.4823	-.5091	-.5758	.1596	.127	.1056	.2334	.2394	.2223
Mean Differences (pvalues)									
25-34 vs 35-49	.0268 (.6421)			.0326 (.1911)			-.006 (.9588)		
25-34 vs 50-64	.0935 (.2410)			.054 (.0046)			.0111 (.5721)		
35-49 vs 50-64	.0667 (.2887)			.0214 (.1292)			.0171 (.4583)		

Vignette A: "Imagine Giulio and Silvia: they are married or cohabiting. Giulio works twice as many hours as Silvia and earns about twice as much. They have no children and no one to help them with the housework.". The prevalent personal value is inside a box. Benjamini-Hochberg adjusted p-values in parenthesis, the p-values refer to a test of equality between age groups in each scenario, these results are not replicated with Wilcoxon rank-sum test .

Table OA9: Decline of the bread-winner model, personal values. Vignette A (Self)

Model	(1)	(2)	(3)	(4)
Dependent Variable	1 if rates Very or Somewhat Appropriate in the equal share scenario, 0 otherwise			
Independent Variables				
Female	-0.030	-0.025	-0.022	-0.031
	(0.0291)	(0.0297)	(0.0298)	(0.0319)
Age Groups (Baseline: 50-64)				
25-34	0.034	0.033	0.021	0.014
	(0.0411)	(0.0417)	(0.0442)	(0.0447)
35-49	0.047	0.050	0.050	0.046
	(0.0325)	(0.0331)	(0.0336)	(0.0338)
Geographical Area of Residence (Baseline: South and Islands)				
North	0.012	0.007	0.006	0.007
	(0.0328)	(0.0349)	(0.0348)	(0.0355)
Centre	-0.065	-0.077	-0.080	-0.081
	(0.0426)	(0.0442)	(0.0441)	(0.0444)
Relocation to a different Geographical Area				
Moved North		-0.003	-0.001	-0.005
		(0.0450)	(0.0449)	(0.0450)
Moved South		-0.018	-0.015	-0.014
		(0.0748)	(0.0747)	(0.0748)
Civil Status (Baseline: Single, Widower, Separated-Divorced)				
Married or Cohabitant			-0.041	-0.041
			(0.0367)	(0.0367)
Having Children			-0.009	-0.006
			(0.0354)	(0.0354)
Framing: Woman Proposing	0.003	-0.004	-0.004	-0.002
	(0.0292)	(0.0297)	(0.0297)	(0.0297)
Controls				
Education/Job				✓
Observations	1501	1443	1443	1443

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Average marginal effects for the probability of rating the equality scenario Very Appropriate or Somewhat Appropriate in the personal values. In columns (2)-(4) we loose 58 observations as we do not have reliable information on the geographical area of birth, or as respondents were born abroad.

Table OA10: Data sources

Variable	description	source
Employment	Fraction of employed women at age and geographical area level	Istat data (downloaded in July 2024).
University degree	Fraction of women with a university degree at age and geographical area level	Own elaboration based on Istat data, "Forze di lavoro – dati trasversali trimestrali " first trimester data (downloaded in July 2024)
High school degree	Fraction of women with a high-school degree (4-5 years) at age and geographical area level	Own elaboration based on Istat data, "Forze di lavoro – dati trasversali trimestrali " first trimester data (downloaded in July 2024)
Childcare	Authorized places for 100 children aged 0-2 years at geographical area level.	Istat data (downloaded in July 2024).
Fraction SA/VA (SoB)	Fraction of male answering Somewhat Appropriate/Very Appropriate as second order belief in Vignette A	Survey data
Fraction SA/VA (FoB)	Fraction of male answering Somewhat Appropriate/Very Appropriate as first order belief in Vignette A	Survey data
Mean Rating (SoB)	Mean rating for males' second order beliefs in Vignette A	Survey data
Mean Rating (FoB)	Mean rating for males' first order beliefs in Vignette A	Survey data