ECONOMIC ANALYSES OF THE IMPACT OF GENDER NORMS ON FEMALE LABOUR SUPPLY

Universität Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften

DISSERTATION

Zur Erlangung der Würde einer Doktorin der
Wirtschafts- und Sozialwissenschaften
doctor rerum politicarum
(gemäß der PromO vom 18. Januar 2017)

vorgelegt von

Luise Görges (M.Sc.)
geboren in Altdöbern

Hamburg, Mai 2018

Vorsitzender:
Prof. Dr. Thomas Siedler, Universität Hamburg

Erstgutachterin:
Prof. Dr. Miriam Beblo, Universität Hamburg

Zweitgutachter:
Prof. Dr. Gerd Mühlheußer, Universität Hamburg

Drittgutachter:
Prof. Dr. Grischa Perino, Universität Hamburg

Datum der Disputation: 11. September 2018

Acknowledgments

This dissertation was written during my time as research and teaching assistant at the Universität Hamburg (UHH). I am indebted to my advisor Miriam Beblo, who ignited my passion for economics, and gently nudged me toward gender topics already during my undergraduate studies. Since then, she has been a role model and mentor who, without fail, advised, encouraged, and supported me. My deepest thanks also go out to Gerd Mühlheußer, who not only spurred my ambition to make this thesis more than purely empirical, but also provided invaluable advice and encouragement.

At UHH, I was fortunate enough to also benefit from the experience of Ulrich Fritsche, Andreas Lange, Grischa Perino, and Thomas Siedler, who took the time to read parts of my work and provided important feedback and guidance. Ingrid Größl, unerringly sensing when it was needed most, dispensed generous amounts of reassurance. Many great colleagues facilitated the writing process both by offering their critical discussions and their friendship: Elisabeth Bublitz, Norma Burow, Leonie Gerhards, Shushanik Margaryan, Eva Markowsky, Kristin Paetz, Annemarie Paul, Melanie Schröder, Artur Tarassow, and many others.

For the more practical aspects of writing this thesis, I thank Claudia Ranft for her kind administrative support, and Leah Bassity for proofreading. The economic experiments I conducted at both the University of Warwick and UHH were an integral part of this research, and I gratefully acknowledge funding from the Department of Economics at Warwick, my department chair Miriam Beblo and the WiSo Graduate School at UHH, as well as the assistance provided by the WiSo research lab team.

Other inputs were also essential to the production of this thesis. It would have not been written without the motivation from and continuous endorsement of my undergraduate mentor and friend Ulf Kadritze, who fostered my love for the social sciences and tirelessly fuelled my desire to learn, read, and write. I also thank my family and friends for their support throughout.

Finally, I thank Arne Pieters. As colleagues and household members, m and f, we have had countless occasions to study the many facets of intra-family bargaining. This dissertation has benefitted tremendously from Arne's generous supply of thoughtful criticism, extensive feedback, and unconditional support.

Contents

General Introduction

> *"The responsibility of married women for child care and other housework has major implications for earnings and occupational differences between men and women, even aside from the effect on the labor force participation of married women. (...) The persistence of these responsibilities may be only a legacy of powerful forces from the past and may disappear or be greatly attenuated in the near future. (...) A person's sex would then no longer be a valid predictor for earnings and household activities. It is still too early to tell how far Western societies will move in this direction."*
> Gary S. Becker, 1991

Motivation

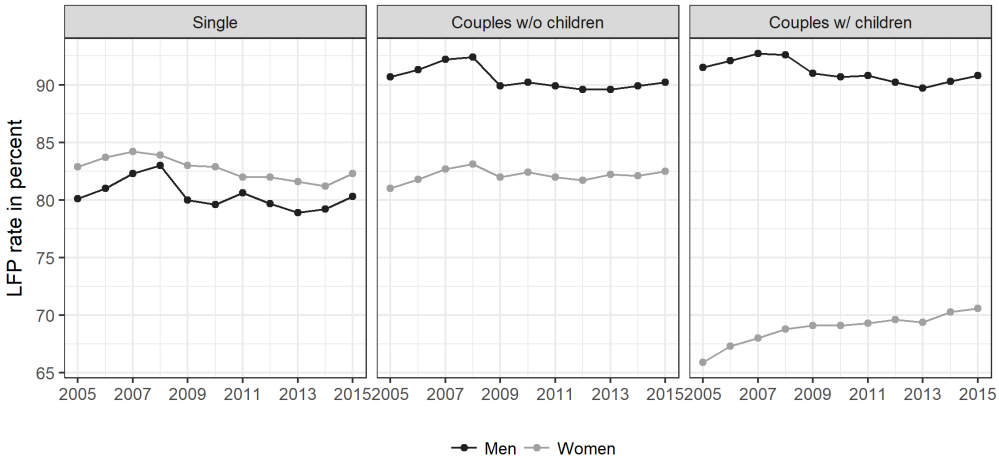
All over the world, economic activity is marked by a distinct gender pattern, and men's and women's social and economic outcomes are, on average, substantially different. Women are less likely to engage in paid market work than men and, if they do, they work fewer hours, different jobs, and earn less. They are underrepresented in more lucrative fields and hierarchy levels, and among business owners and political representatives. At the same time, women are overrepresented among unpaid caretakers: They spend more time in family and house work, even when employed full time, but are also more likely to cut back from paid activities to

care for family members in need (OECD, 2017b). Many gender inequalities exacerbate, and some only come into being, when individuals enter parenthood, which tends to have few ramifications for men's but a profound impact on women's career progression (Kleven et al., 2018; Adda et al., 2017; Lundborg et al., 2017). This asymmetry is closely connected to how parents divide labour and share responsibilites, often choosing mothers as primary caretakers and fathers as primary breadwinners (OECD, 2017b).

Over the past decade, especially in developed countries, remarkably little progress has been made with respect to gendered labour division patterns, as Figure 1 illustrates. Based on data provided by the EU's statistics bureau (Eurostat), it shows the average labour force participation rates for men and women aged 25-49 in the EU-28 over the past decade. While the participation gap is virtually non-existent among singles (left panel) and, at about 10 percentage points, moderately sized among men and women living with a partner but without children (middle panel), it more than doubles among individuals living with a partner and at least one child (right panel). The modest gains women made relative to men in the latter group (about 5 percentage points over the past ten years) may be attributable to policy efforts undertaken by EU members to foster a more equal sharing of responsibilities among family members, especially policies aimed at encouraging fathers to take paternity leave (Beveridge et al., 2016). However, in light of the various gender equality targets, strategies, and action plans set by the EU—and similarly by the OECD and the UN—the improvement appears small. The European Institute for Gender Equality (EIGE), an autonomous body of the EU established to promote gender equality targets, recently lamented "a snail's pace towards gender equality" (EIGE, 2017a).

Policymakers recognise that partners' unequal participation in family and household-related work presents a key obstacle to achieving gender equality in other economic activities. They have also begun to acknowledge the role of gender norms and stereotypes as an important factor impeding the effectiveness of policies aiming

Figure -1-1: Labour force participation rates of adults aged 25-49, living with a partner in the EU-28



Note: Figures based on Eurostat (2017) data, Table '$lfst\_hheredty$'.

to encourage more equal sharing of family responsibilities between men and women (Van Belle, 2016; European Commission, 2016; EIGE, 2017b; OECD, 2017b). In many instances, policies relying on economic incentives alone have done little to shift traditional gender roles toward more progressive family models,[1] but a more nuanced understanding of how they interact with, or may indeed be counteracted by, social norms is lacking. As a contribution toward improving the comprehension of these mechanisms, this dissertation utilizes the versatile methodological tool kit available to economists to compose five essays comprising theoretical, empirical, and experimental analyses of the impact of gender norms on women's participation in the labour market.

[1] One example is the uptake rate of paternity leave, which remains low across the European Union, despite recent policy changes in some countries that aim to raise financial incentives (Van Belle, 2016).

Gender norms and female labour supply in an economic framework

Studying the ways in which gender norms may affect women's labour supply decisions within an economic framework uncovers three main channels: preferences, productivities, and gender identity. To assess these channels and how they relate to gender norms systematically, I augment a simple model of time allocation to market and home production (in the spirit of Reid, 1934; Becker, 1965) with an identity component as proposed by Akerlof and Kranton (2000). The model is fleshed out and analysed more extensively in Chapter 4, but its most basic version presented here serves to illustrate the different channels through which gender norms may influence female labour supply, and thus how the chapters of this thesis are interconnected. For the moment, I abstract from the more complex interactions that need to be accounted for in the presence of a spouse, and simply consider agents as single decision makers. An individual, i, derives utility from the consumption of goods and services that she can purchase in the market, M, and a good that can only be produced in the home, C. Endowed with one unit of time, she can choose how much to allocate to market work; the residual is spent on home production. Market work produces wage income, which in turn allows the purchase of market goods. The level of market consumption, M, is determined by the time devoted to the labour market, t_i, as well as the individual wage productivity, w_i. With regard to time spent outside the market, I consider it an input to the production of child well-being, denoted by C. Similar to the consumption of market goods, utility is increasing in C. The level depends on the time invested in home production, i.e., decreases in t_i, and the individual's productivity in the home, h_i. Thus, agents face a trade-off between production of wage income in the market and child well-being in the home.

$$U_i(M(w_i, t_i), C(h_i, 1 - t_i); I(t_i, t^G)) \tag{1}$$

If we ignore the identity component of the utility function for now, this very basic model already suggests a channel through which gender norms may create the gendered pattern in labour market participation and home production: preferences. If women have a greater relative preference for producing childcare at home than men, it is optimal that they invest more time in the home than in market production. One reason for this may be that producing childcare in the home is not just a means to an end. Margaret G. Reid noted this in 1934: "That much of childcare and training can be delegated is made clear by the presence of nursemaids, nursery schools, and kindergartens. (...) The mother may at times feel that by performing certain tasks, which might be delegated, she is establishing a relationship with her child which is very valuable to her as a mother (...). But the fact that she uses certain tasks as the medium of her contact and does not care to delegate them because they are useful in this way does not mean that they should not be considered productive activities" Reid, 1934: 14f. While much disagreement exists on whether, and to what extent, gender differences in preferences for providing childcare in the home may be rooted biologically (Eagly and Wood, 1999), the aim of this thesis is not to contribute to this debate, but rather to utilize an economic framework to examine systematically the impact of gender norms. In terms of preferences, it is conceivable that norms, at the very least, add to whatever biological differences may or may not exist between sexes to begin with. For example, if gender norms differ across societies, so may the roles men and women are expected to fill. Consequently, spending time in home production may be considered a much more meaningful activity for women in some societies than in others. This hypothesis is the focus of Chapter 1 of this thesis, which investigates the malleability of gender differences in preferences with respect to market work vs. family and own children by social norms and institutions.

While preferences for home production can be considered an indirect channel through which gender norms may enter the utility function of men and women and affect labour supply decisions, another potential channel is more direct. Based on the idea that individuals form a "sense of self" (Akerlof and Kranton, 2000), and "care about who they are" (Bénabou and Tirole, 2011), they may actively seek to behave in line with gender norms. Following Akerlof and Kranton (2000), this can be modeled by allowing for the possibility that individuals not only care about market and home-produced goods, but also about their identity, I, which decreases in the deviation from gender norms, i.e. in the difference between i's time spent in the labour market, t_i, and t^G, the amount of time that is considered socially acceptable for i's gender $G = f, m$. Thus, when social norms dictate $t^f < t^m$, men will supply more labour to the market and spend less time in home production than women, and the discrepancy will, on average, be larger in societies with more traditional gender norms. At the individual level, however, agents may differ with respect to how much value they place on identity relative to the other arguments of their utility function. One mediator for the difference in "attention" individuals pay to gender identity is explored in Chapter 2: gendered language. The essay provides several important insights. First, it presents a theoretical framework to conceptualise how gender identity may impact individual labour supply choices. Second, the empirical application illustrates the challenges that arise when attempting to measure the effect of norms on behaviour—which are difficult to overcome using micro and survey data, and without taking the more complex interactions within the family into account. To better address these issues, the remaining three chapters take a different approach and study the effect of norms on labour division choices from a family economics perspective, by means of economic experiments with real heterosexual couples.

In order to outline Chapters 3–5 within the theoretical framework presented above, suppose a family consists of two adults, m and f, and one or more children, who may affect household decisions but are not considered decision-making

agents. From a theoretical perspective, the household may determine the time allocations of both members in several ways. The first economic accounts of family decision-making assumed partners maximise a joint utility function (Samuelson, 1956) or that a benevolent household head dictates allocations (Becker, 1974b). The optimal consumption level of market and household goods is chosen so as to maximise household (head) utility, given spouses' aggregate time and income constraints. Becker (1973) and Gronau (1973a,b) have shown that household members can increase total output by specializing according to comparative advantage, i.e., assigning the spouse to home-production that imposes lower costs in terms of forgone labour market earnings per unit of output.

The simplistic view on household decision-making as "unitary" can only be reasonable in cases where partners have identical preferences regarding the consumption of market and home-produced goods, and agree on how to share them. In his seminal papers on the theory of marriage, Becker (1973, 1974a) acknowledges that a conflict of interest may arise between spouses when it comes to dividing the gains from marriage, but it was not until later that economists developed precise models of how partners may resolve the tension between shared and conflicting goals. The first Nash bargaining models of family decision-making were presented by Manser and Brown (1980) and McElroy and Horney (1981); Apps and Rees (1988) and Chiappori (1988, 1992) proposed a more general representation of collective bargaining. These models share the assumption that spouses will not miss out on chances for Pareto improvement and that, accordingly, household allocations will be efficient (Browning et al., 2014). A general conclusion is that the same change in wage or income will typically impact household demand differently, depending on which spouse it accrues to. This is because an increase in one spouse's wage or non-labour income does not only expand the household's budget set, but also improves the respective spouse's outside option. If spouses have different preferences, the relative improvement in her bargaining power will shift household demand additionally closer toward her most preferred allocation.

As outlined above, standard collective models without identity suggest a third channel for gender norms to enter the specialisation decision: gender differences in productivity in market and household work, with women being relatively more productive than men in the latter. Norms may amplify gender differences in productivities in various ways.[2] First, they may increase women's *factual* productivity advantage in the home if, based on the parents' gender bias, girls receive more training in household tasks and caretaking than boys. Moreover, they may also elevate *perceived* productivity differences in childcare, as traditional gender norms are typically characterized by the belief that maternal, but not paternal, time is an essential input in the production of child well-being. Gender norms may also amplify gender differences in labour market productivity. One important factor is employer discrimination against women, which is likely to be more pervasive when gender norms are traditional (Givati and Troiano, 2012). Conversely, wage penalties associated with taking parental leave may be much larger for fathers than for mothers when traditional gender norms prevail, e.g., because employers may interpret it as a negative signal for labour productivity in the case of men but not women. Thus, even when partners are equally productive initially, investing the husband's time in home production might impose greater (long-term) opportunity costs on the household than the wife's.

Because it is almost impossible to distinguish the indirect productivity and preference channels from the identity channel using micro or survey data, Chapter 3 studies partners' labour division decisions in a controlled laboratory environment, where gender differences in preferences and productivity are precluded by design. Even in the absence of such differences, couples choose men at a higher rate for wage income production than women. While this result may be interpreted as evidence in support of the identity hypothesis, the analysis does not permit ruling out other channels. Consequently, Chapter 4 builds on this, carefully studying the role of men's over-confidence and women's reluctance to assume sole responsibility

[2] A more thorough discussion of the impact of social norms on these differences is provided in Chapter 4, but the key insights are summarised here.

for family income, and providing the most direct test of the identity hypothesis available in the literature, to the best of my knowledge. Having studied identity and its potential to *reduce* productive efficiency in cooperative households extensively, Chapter 5 finally complements the analysis by exploring the potential of gender norms to *raise* efficiency. The essay moves away from the cooperative framework of family labour division, and studies the effect of gender norms on time allocations in a non-cooperative setting. As such, non-cooperation bears the risk of failing to coordinate on a Pareto-superior outcome, which may be attenuated if gender norms serve as a focal point (Schelling, 1960).

Overview of the thesis chapters and contributions

As illustrated in the previous subsection, this thesis comprises five main chapters. While they share the broader research question of how gender norms affect labour supply and are thus closely interconnected, each chapter investigates more specific sub-questions and is self-contained. Apart from pursuing a common research agenda, another important commonality of these studies is that they combine different subfields of economics, such as labour, family, and experimental economics, as well as applied micro-econometrics, and link to sister disciplines, such as sociology, psychology and linguistics. The chapters use a variety of economic methods, as they build on theoretical, micro-econometric and experimental analyses. In what follows, I briefly describe each chapter and its primary contributions to the economics literature.

Chapter 1 of the thesis is a joint study with Miriam Beblo that investigates the malleability of gender differences in preferences—regarding market work vs. family and own children as well as other aspects of life—by social norms and institutions. The study exploits the German separation and reunification as a natural experiment, and the fact that the two German societies differed markedly in terms of the roles women were expected to fill (workers in the East, homemakers

in the West). Consistent with the hypothesis that gender norms affect preferences, we find that the gender gap is significantly smaller in East Germany than in West Germany. Tracing the evolution of this "gap in the gap" for the aggregate sample and for different cohorts, allows us to identify a causal effect of own experience with specific gender norms and the corresponding institutions. The study makes several important contributions. First, it synthesises the considerable economics literature on gender differences in preferences, and the growing literature on the impact of political regimes on preference formation. Second, by studying real-world preferences using survey data, we expand the literature on gender differences in preferences, which in large part comprises lab studies and thus relies on preference measures for which external validity has not been conclusively established. Third, by providing a comprehensive compilation of historical data sources, we make a substantial effort toward improving the credibility of the identifying assumption of a causal effect of the separation on preferences, i.e., no regional differences before separation—not only for our own study, but also for others who rely on this natural experiment. Finally, our detailed cohort analyses provide unique insights into the mechanism through which gender norms affect preferences, which is primarily through own experience with institutions that support a certain assignment of men and women to specific roles.

Chapter 2 is joint work with Miriam Beblo and Eva Markowsky, and studies a different channel through which gender norms may affect individual labour supply: gendered language. Specifically, we investigate a hypothesis that has recently gained popularity among economists: that speaking a language that requires frequent reference to biological sex causes speakers to behave more gender stereotypically than speakers of gender-neutral languages. We begin by developing an economic model to conceptualise how gendered language may affect labour supply decisions, based on Akerlof and Kranton's (2002) schooling model. We identify key challenges of measuring the direct impact of language on behaviour through cognition and present a systematic review of the literature, which confirms that

the problems we identified have not been addressed adequately. In the empirical part of the paper, we systematically demonstrate that the prevailing approach used in the literature (i.e., studying the behaviour of first- or second-generation immigrants) is unlikely to produce credible causal estimates of the effect of language on gender identity and behaviour. The main contributions of the study are, first, to provide an economic framework to conceptualise the channels through which gendered language may affect decision-making; second, to render a systematic literature overview of existing economic studies on the topic; and third, to supply a systematic empirical assessment that casts doubt on the suitability of the most common empirical approach currently used to identify a causal effect of gendered language on behaviour.

Chapter 3 presents results from a lab experiment comparing real heterosexual couples to randomly matched mixed-sex pairs of strangers. In this real effort experiment, partners play a specialisation game in which they jointly decide on a labour division. Specialisation (one partner performs an unpaid task while the other performs a paid task) is, by design, more efficient than both partners performing the paid task, but it also creates asymmetric access to income. The results show that women are much more likely chosen to perform the unpaid task by real couples than by pairs of strangers. The study makes a number of important contributions. First, it is among the first to study couples' labour division decisions in a lab experiment; as such, the findings provide an important starting point for inquiry into various underlying motives. Second, to the best of my knowledge, it is also the first experiment on specialisation using a real effort task. Other researchers have used real effort tasks to investigate teamwork performing the same, but not specialisation in two different tasks (see Charness and Kuhn, 2011, for an overview). Another set of studies frames their analyses in terms of the gains of marital specialisation, but relies on investment decisions, rather than choices of real effort labour division (Oosterbeek et al., 2003; Güth et al., 2003, 2004; Cochard et al., 2018). Despite its contributions, the study presented in Chapter 3 also suf-

fers from a number of drawbacks: The sample size is small, subjects participating as couples might be differently selected and/or primed than those participating as strangers, and the pay-off asymmetry may induce different strategic motives in strangers than in partners. These provide the motivation for the follow-up study that constitutes the next chapter.

Chapter 4 sets out to examine the motives underlying the gendered pattern in couples' labour division decisions more comprehensively. Among the potential motives under study, the chapter aims to provide a rigorous test of the identity hypothesis, i.e., that partners are willing to sacrifice productive efficiency for conformity to gender norms. The contributions of this study span several dimensions, both theoretical and empirical. It begins by presenting a theoretical model of a cooperative family allocating spousal time to household and market production in the presence of identity concerns. The model is used to distinguish different channels through which gender norms may impact couples' labour division, and carves out the empirical challenges to identifying a direct effect of the desire to conform to gender norms (i.e., identity). Second, the essay presents the first clean experimental test developed to detect an identity effect on couples' labour division. Third, the design also allows the study of two potentially complementary channels: men's over-confidence and women's reluctance to assume sole earner responsibility. While the former has received some attention in the experimental literature with respect to sorting into different compensation schemes (Niederle and Vesterlund, 2007, 2011; Niederle, 2016), this study is the first to show that it may also have considerable effects on intra-family labour division. With regard to women's "responsibility aversion" for family income, the essay is among the first to study this channel. Finally, the study also exploits the lab setting to provide novel insights into the effects of identity on labour supply at the intensive margin, i.e., effort supply.

The final Chapter (5) takes a different perspective on gender norms. While the experimental analyses in Chapters 3 and 4 place a strong emphasis on their

potential to reduce productive efficiency in cooperative households, this chapter uses a lab experiment to explore the possibility that norms improve coordination among non-cooperative spouses, and thereby productive efficiency. Specifically, I vary the presence of gender norms exogenously in a *battle of the sexes* game in which the pure strategy equilibria favour either the male or female partner in terms of pay-off. This is achieved by a neutral labeling of the strategies in the control (A and B) and a specialisation labeling of the strategies (*Career* and *Family*) in the treatment group. The essay shows that gender norms do alter behaviour, particularly for women, but in this case specifically, they fail to raise efficiency of outcomes. The study contributes to the literature in several respects. First, it synthesises the experimental literature on the effiency-raising potential of focal points and the family economic literature on non-cooperative families. Second, it provides novel empirical evidence against the hypothesis that gender norms may be socially beneficial, for they allow improved coordination—an idea that has so far only been studied theoretically. Finally, it provides evidence that the ways in which gender norms affect partners' decision-making are rather complex, which is an important insight, particularly when aiming to design equalising policies to counteract gender norms.

The concluding chapter of the thesis aims to conflate the principal findings of the studies presented in the main chapters and discuss their policy implications. It also provides critical reflections and suggests directions for future research.

Chapter 1

On the nature of nurture. The malleability of gender differences in work preferences[1]

Abstract

We study the malleability of gender-specific preferences for work by exploiting the German division and reunification as a natural experiment. We test whether the two political systems have shaped gender gaps in preferences differentially, based on German-General-Social-Survey data from 1991, 1998 and 2012, an extensive set of register data and historical data from the 19th and early 20th century. Our analyses reveal a substantial East-West difference in the gender gap directly after reunification and no convergence thereafter. A cohort analysis illuminates the mechanism, as the effect is driven by cohorts who grew up during separation, and suggests that institutions, not cultural legacy, are the decisive component.

[1]This chapter is joint work with Miriam Beblo. Valuable comments by Bernd Fitzenberger, Arne Pieters, anonymous referees and the associate editor of JEBO, as well as discussions by participants of EEA 2014, EALE 2014, SOLE 2015, IAFFE 2015, Annual Meeting of the German Verein für Socialpolitik 2015, ASSA Meeting 2016 and research seminars at the IAB, IZA, University of Paris (Pantheon-Sorbonne) and Humboldt University are gratefully acknowledged. Accepted for publication in the *Journal of Economic Behavior & Organization*.

1.1 Introduction

Despite the enormous progress toward gender equality in most Western societies over the past several decades, women still earn less than men on average, are less likely to be active in the labour market, work in different jobs, supply fewer hours of work, and are more likely to interrupt their employment for child-rearing or to provide other family-related services (OECD, 2017b). Recently, economists have devoted much attention to preferences as an important driver for these differences in economic outcomes. Over the past 20 years, they have produced a battery of empirical studies on gender differences in preferences for risk, competition and regarding others (for an overview, see Croson and Gneezy, 2009; Bertrand, 2011).[2] While the preference measures in question stem mostly from artificial laboratory settings, recent evidence shows that they are in fact quite strong predictors for real world decisions that directly affect labour market outcomes (Buser et al., 2014; Reuben et al., 2017). Furthermore, recent studies demonstrate that a sizable part of the gender earnings gap among US college graduates is explained by preferences regarding career, family, and job attributes (Grove et al., 2011; Wiswall and Zafar, 2017).

Against this background, a better understanding of how gender differences in preferences evolve seems essential in order to devise policies promoting gender equality in labour market outcomes. The effectiveness of available strategies likely hinges on the malleability of men's and women's tastes for career vs. family work. It is for this reason that we investigate the "nature of nurture", i.e., the impact of policy on gender differences in work preferences in the context of Germany.[3] The division of the country after WWII into distinct political systems, and its

[2] See also Nelson (2015) and Filippin and Crosetto (2016) for a critical assessment of magnitude and economic relevance of gender differences in preferences for risk.

[3] Our goal is not to get to the bottom of the nature *versus* nurture debate, as our analyses do not allow drawing conclusions regarding a potential biological foundation of gender differences in preferences, nor its precise interaction with the societal environment. Instead, we investigate the social structure mechanism, which posits that gender differences in individual behaviour follow in part the prevailing social structure, i.e., the way societies allocate men and women into different

reunification in 1990, provide a natural experiment to study the potential for policy to shape preferences using relatively short-term interventions.

Notwithstanding the rapidly growing number of experimental studies on gender differences in preferences, surprisingly little effort has been made to understand their root causes. Only a handful of field experiments provide valuable insights. Gneezy et al. (2009) study the role of culture and social structure by comparing the gender differences in competitiveness across a patriarchal tribe in Tanzania and a matrilineal tribe in India, and find that the gender gap is reversed in the latter. Booth and Nolen (2012a,b) examine gender differences in competitiveness and risk behaviour across school types (mixed-sex versus single-sex schools) and detect a gender gap only among mixed-sex, but not among single-sex school students.[4] Bertrand (2011) raises concerns about the "evolutionary distance" between the societies compared by Gneezy et al. (2009), and, in the case of the Booth and Nolen (2012a,b) experiments, about selection into the different school types. Both concerns are less pressing in our study of Germany: We compare two societies of presumably minimal evolutionary distance since East and West Germans share a common past and cultural identity up to the artificially imposed separation after WWII. Moreover, a "selection" of individuals into the different Germanies did not occur, at least at the time of the separation.[5] Our study is the first that builds

roles (Eagly and Wood, 1999), which by itself is determined both by institutions and cultural legacy.

[4]See also Dreber et al. (2011); Cárdenas et al. (2012) and Sutter and Glätzle-Rützler (2015) who study gender differences in the competitive behaviour in children and adolescents and document the importance of socialization and social learning, as well as the influence of culture. Säve-Söderbergh and Lindquist (2017) also find evidence that gender differences in risk-taking might not be present among children aged 10.

[5]We explore the validity of our assumption on minimal evolutionary distance in Section 1.3, and alleviate concerns that cross-migration flows between the two Germanies could pose a threat for our identification in Section 1.2 because of the way our sample is constructed: We only include individuals born after 1941, who were too young to self-select into migrating before 1954, when exiting the GDR without a departure permit was criminalized. In the robustness checks in Section 1.6, we repeat parts of our analyses exploiting information on where respondents lived when they were aged 15, and thus when they were presumably too young to have self-selected into migration.

on a natural experiment and real-world preference measures to analyse the role of culture and institutions in gender-specific preference formation. Notably, we are not the first to suspect that culture and institutions affect female labour market outcomes. To the extent that the outcomes studied by Alesina et al. (2013); Fernández (2013); Fogli and Veldkamp (2011); Fernandez and Fogli (2009); Fortin (2005) (e.g. labour force participation) represent revealed preferences, both culture and institutions have indeed been shown to be influential factors in the formation of gender-specific preferences. What remains an open question, however, is whether their influence evolves in an almost evolutionary process over a long period of time, or whether political interventions can have a short-term impact and gender differences in preferences are malleable.

The German separation and reunification has attracted the interest of several economists who sought to identify the causal impact of differential political regimes on various preference and attitude variables, such as tax morale (Torgler, 2003), preferences for redistribution (Alesina and Fuchs-Schündeln, 2007), trust in others and government institutions (Rainer and Siedler, 2009), gender role attitudes (Bauernschuster and Rainer, 2012), inequality perceptions and equity norms (Kuhn and Shen, 2013), conspicuous consumption (Friehe and Mechtel, 2014), attitudes toward work (Campa and Serafinelli, 2016), or behavioural variables such as college attainment and labour market outcomes (Fuchs-Schündeln and Masella, 2016). Our study extends this literature by providing the first analysis of the impact of political regimes on *gender differences* in preferences. Friehe and Mechtel (2014) also note an overall gender gap in conspicuous consumption, but do not study to what extent it was shaped differentially by the two political regimes. Most closely related to this chapter, the study by Bauernschuster and Rainer (2012) provides us with important insights regarding the cultural norms and attitudes toward working mothers and wives in the Eastern and Western parts of Germany. However, their focus is not on gender differences in these attitudes, and they examine neither the effect of length of exposure to GDR institutions nor

heterogeneous developments over time in different cohorts. Adler and Brayfield (1997) find a difference in East and West German women's attitudes toward work in 1991, and Campa and Serafinelli (2016) can attribute this effect to employment experience in the GDR. Since both studies focus on women, neither allows drawing conclusions on the differential evolution of the gender gap in preferences for work over time.

Our contribution lies in a synopsis of the experimental research on gender differences in preferences and the mostly survey-based literature that uses the separation and reunification of Germany to assess the effect of political institutions on preference formation. Using the same natural experiment, we study the effect on *gender-specific* preferences. During separation, which lasted for 41 years, the political systems in East and West Germany differed markedly in their institutions and the role for women in society they promoted. While in East Germany women were expected to participate in the labour market to the same extent as men, the West German society fostered a much stronger sorting along gender lines into separate spheres (home and market), so male and female experiences differed more strongly in the West (Rosenfeld et al., 2004). Thus, if preferences were malleable through social policy and institutions over a relatively short time span, we should find different gender gaps in preferences for work in East and West around reunification. Following the same reasoning, gender differences in preferences may have become more similar across the two regions as time progressed because Western institutions were quickly established in the East, and gender-specific assignments most likely assimilated in reunified Germany. Observing a convergence following the political integration, over yet a shorter period of time (we observe roughly 20 years until 2012), would thus reinforce the case for the power of institutions and the malleability of gender-specific preferences. Disentangling the dynamics into cohort, time or life cycle effects, allows us to shed light on the mechanisms[6].

[6]These analyses also serve as a robustness check, supporting our interpretation of regional differences in gender-specific preferences at the end of the German separation as causal to the

Our analysis allows several important conclusions regarding "the nature of nurture": In 1991, even though we detect a significant gender gap in preferences for work in both parts of the country (with women finding work a less important aspect of life than men), it is significantly smaller in the former GDR. That is, we find that gender differences in preferences at the aggregate are malleable within a very short time span (less than 50 years). By 2012, the gender gap in the West has fallen considerably but remained statistically significant, whereas the gap between Eastern men and women has vanished completely. Thus, in both regions the gaps seem to follow a generally fading trend, as individual attitudes become more progressive and work preferences less gender-typical. Because this trend took off earlier in the former GDR, East German regions maintain a head start such that the "gap in the gap" (we will refer to this as the "GiG") across the two parts of the country remains economically and statistically significant at almost the same level, even 20 years after reunification. While at first glance this may suggest that institutions' short-term effectiveness in shaping gender differences in preferences is rather limited, a detailed cohort analysis reveals that the effect is driven entirely by cohorts who spent at least 15 years of their life in the East before the fall of the wall in 1989, i.e., have acquired the critical dose of experience with GDR institutions.

These findings underline the power of a short-term political intervention for two reasons: They suggest that the effect we find in the aggregate sample is indeed causal to the differential experiences acquired within the two political systems during separation and not driven by historical differences. They also reveal that institutions, not cultural legacy, are the decisive mechanism because the effect is strongest among those who have experienced employment and institutions in the GDR for a longer time. These results are robust to the inclusion of a broad set of individual and macro-level control variables and to a series of further robustness checks: We verify that our results are not sensitive to the inclusion of

social structure experienced in the GDR, because only then would we expect them to fade out in younger cohorts. This is indeed what we find.

pre-separation female labour force participation rates using historical data from 1934; we examine whether selection effects drive our results by categorizing respondents based on the region they lived in during their adolescence (rather than using their residence at the time of the interview) and are thus also able to study East-West migrants; we analyse potential heterogeneity across the East German federal states and conduct a placebo test between northern and southern German federal states.

The chapter is organized as follows: Section 1.2 briefly reviews the division of Germany into two countries after WWII, the respective political contexts of female employment during and after separation and, against this background, reflects theoretically on the gender gaps in work preferences in East and West. Section 1.3 investigates historical data from the late 19th and the early 20th century to validate our assumption that differences between the two regions prior to the division were not systematic. Section 1.4 introduces the data sources and preference measure we use for our main analyses, and outlines our empirical strategy. Section 1.5 supplies the regression results for the aggregate sample and separate analyses by cohorts. Section 1.6 explores the robustness of our results to various potential selection and heterogeneity issues pre and post separation. Finally, Section 1.7 offers a discussion of the results and concludes the chapter.

1.2 The German separation and reunification and gender differences in work preferences: What do we expect?

The German separation offers an ideal natural experiment to study the malleability of gender-specific preferences for work because it allows us to examine the effect of the differing social structures in East and West Germany that were imposed exogenously after World War II on an evolutionary uniform population. While

we provide support for this claim by evaluating historical data in Section 1.3, this section is intended to motivate our considerations as to why we may expect differences across regions, laying out the different institutional backgrounds provided by the two regimes. It will also explain why the self-selection of individuals into the two Germanies *during* the time of separation is not a concern for the present study.

After World War II Germany was divided into two distinct countries along the Soviet occupation zone borders. Having shared a common cultural past as one country until then, the German Democratic Republic (GDR) was constituted on the grounds of the Soviet occupation zone, which covered the five Eastern Laender. The remaining 11 Länder, occupied by the Americans, British, and French, formed the Federal Republic of Germany (FRG). In 1989, a peaceful revolution led to the fall of the Berlin Wall and a swift political reunification of the two German parts soon followed in 1990 (with a rapid imposition of monetary union and FRG institutions in East Germany; see Krueger and Pischke, 1995).

During the political division, people living in the two German states received differential treatment through labour market and educational institutions, as well as gender role norms, particularly with respect to female employment. Migration flows from the FRG (West) to the GDR (East) were practically negligible. Migration in the reverse direction was in principle possible until 1961 (Fassmann and Munz, 1994), but exiting the GDR without a departure permit and handing in one's ID card was criminalized after 1954 (§213 criminal code of the GDR StGB-DDR, 1968). After the construction of the wall and until its fall, from 1961-1989, only about 800,000 GDR citizens managed to legally depart to the FRG (Fassmann and Munz, 1994). Since we will focus on individuals born 1942 or later, and exploit information on where respondents lived when they were 15, selective migration should hardly be an issue for our analyses.[7]

[7]Moreover, as detailed in the following paragraph, differences in female labour force participation rates across the two German states did not accelerate until the 1960s, which makes it

From earlier studies we know that female labour force participation (LFP) was much higher in the GDR (East) than in the FRG (West) (Holst and Schupp, 2001; Rosenfeld et al., 2004; Bothfeld, 2005). Starting at about 8 percentage points, the difference accelerated from the 1970s, when female LFP started rising above 80 percent in the GDR, hitting 90 percent in 1989, whereas in the FRG it only began exceeding 50 percent in the 1980s. This can be seen in Table 1.1, which summarizes male and female labour market and family behaviour in East and West Germany from 1950 to 2010. Consequently, the East-West difference in the gender gap in LFP was small in the early years of the separation. It started becoming sizeable in the 1960s and 1970s, and reached a maximum in its last measurement preceding re-unification, where the gender gap had fallen to 2.4 percentage points in East Germany, less than 10% of the West German gender gap of 26.7 percentage points at the time.

The differential development in the gender gap in LFP coincided with pronounced institutional differences between the two Germanies. In the FRG in the 1950s and 1960s, many social and tax provisions were introduced that favored the breadwinner household with nonworking spouse, such as joint taxation of married couples (Gerhard, 1992). For mothers it was almost impossible to work full time, and even part-time employment was difficult to organize, because up until the 1990s, child care for pre-schoolers was scarce and elementary schools had varying daily schedules or would even close over the lunch hour (Ostner, 1993). The GDR, on the contrary, emphasized the equality of spouses in the Family Law Code (Familiengesetzbuch) in 1965 and enforced women's obligation to work equally to men's. Due to the state provision of universal child care and the East German citizen rights based on the status of labour force workers, most women, including mothers, were employed full-time (Duggan, 1995). In fact, the criminal code of the GDR classified the avoidance of work as anti-social behaviour and a criminal act that could be prosecuted and punished with up to 5 years in prison (§249 StGB-

even more unlikely that individuals, women in particular, selected into a specific regime based on their work preferences.

DDR, 1968). Consequently, the state supported maternal employment. In 1950, the Mother and Child Care and Women's Rights Acts (Gesetz über den Mutter- und Kindesschutz und die Rechte der Frau) established "a network of public child care centers, kindergartens, and facilities for free school meals, maternity leave, and days off to care for sick children" (Cooke, 2006: 5).

Similar to female LFP, fertility outcomes also only began to diverge *during* the years of separation. In Table 1.1 we see that the number of marriages and births per 1000 inhabitants was very similar in both countries until the 1980s, when the GDR regime in the East started to incentivize both, e.g. through prioritization of married or single parents with children in housing construction and allocation (Kröhnert and Skipper, 2010). The drastic reversal of these differences shortly after the collapse of the GDR in 1991 may partly reflect a timing effect, but it may also be attributed both to the cancellation of the GDR incentives and to the economic uncertainty of those years (Chevalier and Marie, 2017). Both marriage and fertility behaviour converged until 2010.[8]

Given the contrasting roles that the two states promoted for women and that women in both states seem to have lived up to, we expect the gender *differences* in preferences to *differ*, if they are at all affected by social structure. Provided that we find such a regional gap in the gender gap (GiG) between East and West directly after reunification, the unification process may serve to study its dynamics. However, although women and men were allocated to the role of workers with almost equal probability in the GDR, this did not necessarily extend to equal representation across occupations and hierarchy levels. As a matter of fact, even though the share of women who attended professional colleges and universities was much higher in the East than the West, East German women only entered

[8]It has to be noted that there were also institutional differences with respect to the available family planning technology, in line with the differential roles for women in society promoted by the two regimes. During separation, East German women had easier access to the contraceptive pill and used it at a higher rate than their West German counterparts (Starke and Visser, 1994). Abortions could be performed legally until the third month of the pregnancy (David, 1992), while this was only possible in the West if very strict criteria were met (Robertson, 2004).

Table 1.1: Labour market and family outcomes during and after separation

		LFP rates*			Family outcomes**	
		Male	**Female**	**Difference**	**Births**	**Marriages**
1950	**East**	95.5	52.4	43.1	16.7	11.7
	West	92.9	44.2	48.7	16.1	10.6
1960	**East**	96.7	66.5	30.2	17.04	9.7
	West	91.4	48	43.4	17.4	9.5
1970	**East**	97.2	81.8	15.4	13.88	7.7
	West	88.3	46.3	42.0	13.3	7.4
1980	**East**	94	86.9	7.1	14.64	8.0
	West	84.4	50.2	34.2	10.1	6.3
1989	**East**	93.3	90.9	2.4	12.11	7.9
	West	82.7	58.5	24.2	10.9	6.5
1991	**East**	84.2	77.2	7.0	7.22	3.8
	West	81.1	60.7	20.4	11.24	5.6
2000	**East**	77.9	69.3	8.6	7.6	4.3
	West	80.7	62.9	17.8	9.78	4.7
2010	**East**	82.2	76.0	6.2	8.31	4.5
	West	82.3	70.7	11.6	8.29	4.7

Note: Figures for East including, West excluding Berlin.
*) The labour force participation rate is the labour force as a percent of the working age population.
**) Family outcomes are the number of incidents per 1000 inhabitants.
Sources:
1) LFP East, 1950-1989: BMAS (Federal Ministry for Labor and Social Affairs Germany), 2006: p. 144, Table 3.1.1.1.
2) LFP East and West, 1991-2010: Destatis (Federal Statistical Office Germany), 2014a:, p. 137, Table 5.4.
3) LFP West, 1950 -1970: Own calculations based on Destatis (Federal Statistical Office Germany) (1972). Calculated as the share of people in the workforce younger than 65 (Table 1, p. 140) divided by the total population aged 15-65 (Table 6, p. 95). Note: Values for 1960 are from 1961.
4) LFP West, 1980: Destatis (Federal Statistical Office Germany), 1982: p. 94, Table 6.2.
5) LFP West, 1989: Own calculations based on Destatis (Federal Statistical Office Germany) (1990). Calculated as the number of workforce participants aged 15-65 in 1989 (Table 6.5, p. 96) divided by the number of people aged 15-65 in 1988 (Table 3.10, p. 54).
6) Births East, 1950-2010:Destatis (Federal Statistical Office Germany), 2016: p. 7, Table 1.6.
7) Births West, 1950-2010: Destatis (Federal Statistical Office Germany), 2016: p. 6, Table 1.5.
8) Marriages East: Destatis (Federal Statistical Office Germany), 2016: p. 7, Table 1.6.
9) Marriages West: Destatis (Federal Statistical Office Germany), 2016: p. 6, Table 1.5.

into 16 traditionally female vocational tracks out of many hundreds available to them (Nickel, 1992, cited by Cooke, 2006). Gender-specific job segregation was not less pronounced in the GDR (Rosenfeld and Trappe, 2002), and still looks similar today (Beblo et al., 2008).[9] If individuals' valuation of work is related to the type of job they are engaged in, there may be systematic gender differences in these valuations. Thus, we expect gender differences in preferences for work to be *smaller* in the East, not necessarily non-existent.

If, instead, the GDR regime had enforced high female labour force participation against the true preferences of women, we should observe *equally large* gender differences in preferences for work in both parts of the country. The absence of a GiG does not seem entirely implausible given the heterogeneous labour market developments in male and female employment after reunification documented in Table 1.1. Differences between the regions narrow, both because the LFP gender gap widens in the East and decreases in the West. Hunt (2002) finds a large drop in East German female employment by 23 percentage points over the four years following reunification, compared to a smaller drop of 17 percentage points for men.[10] This would be consistent, she notes, with both a supply side and a demand side explanation: On the supply side, a convergence in female preferences for home production could have been the cause for the disproportionate female exit from the labour market. If this explanation were supported by our data, we should find equally large gender gaps in preferences for work around 1991, which would

[9]Becker (1985), even before the fall of the Iron Curtain, notes that women's higher integration into the labour force in socialist countries is usually not accompanied by a reduction in their housework and childcare obligations. He remains agnostic with respect to the question why societies, even under socialism, seem to assign these reproductive responsibilities primarily to women – the reasons may include women's intrinsic comparative advantages in home production as well as their exploitation (Becker, 1985:S40f.). But the fact that this has been the common practice in most societies, he argues, forced women to supply less energy per time unit devoted to market work, and thus explains occupational gender segregation and differences in pay.

[10]This development was accompanied by a similar development of the gender pay gap. From comparable levels of about 25% at the time of reunification (Krueger and Pischke, 1995), the East German pay gap dropped to 8% in 2013 (Destatis (Federal Statistical Office Germany), 2014b).

indicate a relatively weaker impact of short term social policy interventions. On the demand side, the large drop in female employment would be consistent with a convergence in employers' taste for discrimination, with East German women being affected most strongly. In that case East German women's preferences for work should differ less from men's initially, but could be responsive to employer discrimination in the following years.

If the exposure to distinctive institutions in GDR and FRG did affect gender differences in preferences, various *dynamics* in the GiG in the years after reunification might be plausible. The empirical pattern may shed light on the question which channel is more important in shaping preferences, cultural legacy or institutions. As the most plausible pattern we may anticipate that the gender gaps in the East and in the West converge toward the same level on several grounds, the most obvious being that the whole country is now governed by West German institutions. Whether gender gaps in preferences converge towards Western levels, or the other way round, will be left to empirical testing. There are good reasons why we could expect to observe an East-to-West convergence: Despite the GDR state's progressivism in terms of the gender roles it promoted, the legislation delegated a large share of family-related obligations into women's realm of responsibility, as exclusively married women had a monthly day off to perform housework, and only mothers had fewer weekly working hours and were eligible for parental leave (Duggan, 1995). After reunification, because these supporting policies were no longer part of the institutional framework, East German women's preferences may converge to those of the West German women, who always had to balance work and family responsibilities on their own account, while men, in their role as breadwinners, were responsible for providing income. Other considerations suggest a West-to-East convergence: West German institutions, despite still promoting a modified breadwinner model through the taxation system, have become more progressive in supporting reconcilability of motherhood and labour force participation (Rosenfeld et al., 2004).

The *pace* of the adaptation process depends crucially on heterogeneous convergence dynamics for differentially affected individuals that may well be masked by the aggregate effect. We therefore distinguish the general time effect from cohort and life cycle effects to unveil the dynamics between and within different birth cohorts in our sample. The *cohort effect* refers to the size of the GiG for different birth cohorts at a given point in time, while the *life cycle effect* describes the development of the GiG for a given cohort at different points in time. As for the cohort effect, two distinct patterns are plausible. If nurture works primarily through intergenerational transmission (or legacy) of culture, the GiG is predicted to remain of similar size across cohorts. If institutions are the main channel, on the contrary, it ought to become smaller the younger the cohort, since age is correlated with length of exposure. According to the *life cycle effect*, we expect the GiG to follow an inverse U-shape, first becoming larger with progressing time and entering child-bearing age and then decreasing again as individuals reach the middle-age life phase when the importance of work diminishes in anticipation of retirement (Helson and Soto, 2005; Ekerdt et al., 2000; Kalleberg and Loscocco, 1983). If gender differences in preferences are shaped by the different experiences men and women accumulate in a specific society the increase in the GiG in childbearing age shall be particularly marked in older cohorts. The older a cohort at the time of reunification, the more likely a given woman will continue to fill the role that was routinely chosen in her "native" society (i.e., as worker in the GDR, or homemaker in the FRG). Thus, if the institutions constitute an important factor in shaping gender differences in preferences, we expect that any life cycle patterns shall be different in East and West, with West gender differences in older cohorts increasingly exceeding East gender differences over time. At the aggregate, however, the cohort and the life cycle effects may potentially cancel out and produce a zero time effect. For this reason, a cohort analysis seems indispensable to uncover the underlying dynamics and conclude on how precisely nurture shapes gender differences in preferences.

1.3 Pre-separation Germany: Minimal evolutionary distance assumption

The preceding Section 1.2 laid out our theoretical considerations regarding the impact of political regimes on gender differences in work preferences, assuming "minimal evolutionary distance" between the East and West German societies. Before turning to the main analyses, we examine several historical data sources from the 19th and early 20th century in order to validate this assumption. Our aim is to assess whether there were systematic differences in outcomes that could be related to differential gender gaps in work preferences *prior* to separation. To this end, we draw on the ifo Prussian Economic History Database (iPEHD), which contains detailed information on agricultural, industrial and occupational structure, educational systems, and demographic structure at the district level in the second half of the 19th century (for a comprehensive description of the data set, see Becker et al., 2012). These historical data are available at the district level (335 in total) for several years during the 19th century. We augment the relevant indicators with statistics from the yearbook of the Statistisches Reichsamt (1936), which includes 1930s data on industry sectors, labour force participation, marriage, and fertility behaviour.

Using both data sources, we identify districts that, in succession of WWII, became part of the GDR and the FRG, in order to determine whether systematic structural differences already existed in the late 19th and early 20th century. It must be noted that about two thirds of the GDR can be mapped to Prussian districts, whereas about half of West Germany falls within Prussia, leaving mostly the South outside the borders. A map illustrating the match is provided in Appendix 1.8 (see Figure 1-3).[11] With respect to our second historical data source, the 1936 Yearbook of the Statistisches Reichsamt, not all districts could be unambiguously

[11] From the 335 Prussian districts, we were able to assign 198 to either FRG or GDR territory, the rest falls outside the post-war German borders.

sorted into GDR or FRG territory due to overlaps or regions that, after 1945, were no longer part of Germany. Nonetheless, the coverage is still well above 80 percent.

Table 1.2: Socio-economic indicators in Eastern and Western German regions, pre-separation

	1849*		**1882/86***		**1933/34****	
	East	**West**	**East**	**West**	**East**	**West**
Employment by sector						
Agriculture %	72.76	74.46	56.32	49.91	18.40	20.24
Handcraft %	12.83	12.83	—	—	—	—
Industry %	6.78	6.88	26.54	31.61	—	—
Industry and Handcraft %	—	—	—	—	42.53	37.93
Services %	7.63	5.83	12.37	12.83	9.06	9.03
Retail %	—	—	4.78	5.66	—	—
Retail and Transport %	—	—	—	—	15.93	16.08
Free occ./Self-employed %	—	—	—	—	14.08	12.16
Total workforce (in million)	2.48	2.15	2.13	1.77	14.15	34.31
Female share of employees %	—	—	—	—	34.32	31.28
Girls' share elementary school %	49.37	48.68	50.11	49.51		
Marriages per 1000 inhabitants	—	—	—	—	9.85	8.97
Births per 1000 inhabitants	—	—	—	—	14.40	14.58
Child-woman ratio %***	64.43	64.18	—	—	—	—

Sources: Own calculations based on Prussian data sets of 1849, 1882 and 1886 (Becker et al., 2012) and on Statistisches Reichsamt, 1936: 27, 37, 306 for 1933/34.
*) Only Prussian districts within the later GDR and FRG boundaries (1948 to 1989). 1882: Total workforce without handcraft.
**) All regions of the later GDR boundaries, including Berlin, and FRG boundaries, excluding Berlin (1948 to 1989).
***) The child-woman ratio is calculated as the number of children under the age of 5 per women aged 15-45.

Table 1.2 summarizes the indicators related to our research question compiled from these sources. It begins by listing the shares of employees in economic sectors in East and West districts for the years 1849, 1882, and 1933. The general trend is that agriculture has declined in relative workforce (from three-fourths to around one-fifth), while the industry sector has gained (from below 7% to around 40%, including handcraft). Services have increased only slightly in importance; retail is

first mentioned in 1882, while transport appears as a sector in 1933. Differences between East and West regions seem to evolve in the second half of the 19th century due to a faster industrialization process in the West, which then reverses and partly converges until 1933, as documented by the respective shares of the agricultural and industry sectors in 1882 and 1933. All in all, structural differences in types of economic activity do not appear to vary in a systematic manner between the East and the West prior to the political separation that would alter our interpretation of observed differences thereafter. If any, East German women would have started with *lower* labour force participation into separation, as female involvement in the industry sector has traditionally been lower than in agriculture (Goldin, 1995). We might expect this to bias any potential regional differences in gender differences toward zero, but the most important sector for female employment, the service sector, does not seem to differ in size across regions since the 1880s.

Regarding the link between Protestantism, girls' education, female literacy, and economic outcomes throughout Prussia established by Becker and Woessmann (2008), we also examined gender-specific school enrolment and literacy. Table 1.2 shows that in the years 1849 and 1886, about 50% of elementary school pupils were girls, both in the East and the West German county average. We do not see any systematic differences here, neither for male nor female literacy rates (only available for 1871, hence not displayed).

The percentage of women among all employees averaged to about one third in pre-WW II Germany, varying between 26% and 38% across regions (Landesarbeitsamtsbezirke). Saxony (East) and Bavaria (West) showed over-proportional and Westphalia (West) and Thuringia (East) under-proportional female labour force participation which resulted in only a marginal difference between the historical halves of Germany (on average, 34% of women in Eastern regions were employed as compared to 31% in the West).

With regard to the demographic past, the Prussian data provide numbers on population-age groups from which we derived the child-woman ratio for East and

West districts. The child-woman ratio gives the number of children up to age 5 divided by the number of women of child-bearing age (15-45). We calculated 64% for both East and West (matching exactly the average level documented by Becker et al., 2013). The statistics by the Statistisches Reichsamt (1936) further document similar marriage and fertility behaviour between later GDR and FRG districts. In 1933, marriage distributions look very much alike between provinces later forming the GDR and those forming the FRG. The number of marriages per 1,000 inhabitants averaged 9.85 (East) versus 8.97 (West). The number of births differs even less between the East and West, counting 14.40 births per 1,000 inhabitants in the Eastern provinces compared to 14.58 in the West. To summarize, our data seem to underpin a similar marriage and absolute fertility behaviour across Germany.[12] All in all, the comparison of pre-separation labour market and family outcomes across regions provides support for the "minimal evolutionary distance" assumption, which our identification strategy relies on. There is little evidence supporting the alternative explanation, that differences in preferences for work, if they exist, have anteceded separation.

[12]Non-marital fertility, however, is not documented in these data sources. We know that in the late 19th century, non-marital fertility was about twice the level in areas that would later become the German Democratic Republic than in those that would become West Germany, and is still higher today (Klüsener and Goldstein, 2014). To the extent that non-marital family formation is linked to work preferences, this may arguably weaken the notion of minimal pre-separation differences. We would then be unable to attribute any gap in the gap, or GiG, in preferences we observe right after the separation years to individuals' differential experiences acquired within the two states. Since, according to Klüsener and Goldstein (2014), Bavaria had similarly large non-marital fertility rates as the Eastern German regions in the 19th century, we are able to investigate this issue in a sensitivity analysis. We will discuss the results in Section 1.5, they reveal that the East German gender gap still significantly undercuts the one found in Bavaria. We thus feel confident concluding that this potential threat does not invalidate our identification strategy so that we can proceed with our primary analysis.

1.4 Methodology

1.4.1 Data & sample

To study whether the GDR institutions had an effect on gender differences in work preferences, and whether the reunification and the subsequent mixture of West institutions and East legacy further affected a potential regional heterogeneity, we combine data from the German General Social Survey (ALLBUS) with official German register data. The ALLBUS regularly surveys a random sample of the German population on a wide variety of social and political topics, as well as demographic background characteristics. It is comprised of independent cross-sections, i.e., respondents are not surveyed repeatedly, which unfortunately prohibits the use of panel methods. The survey began with West German inhabitants in 1980 and has included East German respondents since 1991 (Terwey, 2000). For our research design, we use three cross-sections (1991, 1998, and 2012) since only these include our dependent variable, and they allow us to cover a meaningful time horizon from just after reunification up to two decades later. Additionally, in two of these waves respondents provide information on the federal state they lived in when aged 15 (as opposed to where they lived at the time of the survey), which allows us to investigate selective migration issues in the sensitivity analyses.

To reduce the noise potentially introduced by individuals with more heterogeneous cultural backgrounds we restrict our sample to respondents of German citizenship. In doing so, we are able to avoid any issues of selective migration from outside Germany to either East or West that relates to work preferences. Furthermore, we exclude individuals above the age of 50 to avoid issues related to early retirement policies, a strategy the German government adopted extensively in order to mitigate unemployment during the restructuring of the East German economy after the formation of the monetary union (Krueger and Pischke, 1995).[13]

[13]Moreover, there is evidence from the sociology and psychology literature that individuals' self-concept, i.e., their assessment of aspects they consider important in life, changes substantially in the middle-age life phase (Helson and Soto, 2005). One's assessment of the importance of work

Finally, by excluding people born before 1941 and using the information on residence during *adolescence* for the remaining respondents, we can alleviate selection concerns relating to the arguably greater migration opportunities for East Germans before the GDR regime tightened its departure regulations in the 1950s.

We complement the survey information provided by ALLBUS with official register data compiled from different sources in order to construct a comprehensive set of federal-state-level macro-controls (Destatis, 2015). We will provide further details on the controls in Section 1.4.3.

1.4.2 Variables

Main independent variable

The key estimator in our set-up is a dummy variable indicating whether a respondent i lives in one of the former GDR federal states. Thus, the dummy $East_i$ takes on the value 1 if the respondent is a resident of the Eastern part of Germany (the set E) at the time of the interview. The dummy is 0 if the respondent resides in the Western part of Germany.

$$East_i = \begin{cases} 1 & \forall i \in E = \{ \text{ East Berlin, Brandenburg, Mecklenburg Pomerania, Saxony, Saxony-Anhalt, Thuringia} \} \\ 0 & \forall i \notin E \end{cases}$$

As a sensitivity analysis, we repeat all analyses excluding Berlin residents (East *and* West) and the results are very similar. Moreover, we use a refinement of the $East_i$ variable for the robustness checks in Section 1.6. In two of our cross-sections (1991 and 2012), respondents provide information on the German federal states they predominantly resided in throughout their adolescence. This variable thus takes on the value 1 if individuals report having spent their youth in one of the Eastern states and zero if in one of the Western states. Assigning the

and job aspects also seems to change drastically in this phase of life, as retirement grows closer (Kalleberg and Loscocco, 1983; Ekerdt et al., 2000).

region in which a respondent spent her adolescence, rather than the region of her present residence, may enable us to assess the importance of socialization, and the influence of political environments during that critical period for preference formation more precisely. This strikes us as a particularly promising purpose, as almost all previous studies on the impact of differential political regimes on various preference and attitude variables were forced to exclusively rely on the region of current residence, and could not exploit the important additional information on residence during adolescence.

Dependent variable: work preferences

We rely on an ALLBUS item that surveys respondents' ratings of the importance of "job and work" as one of several "spheres of life" as our dependent variable. Importance ratings are given on a 7-point Likert-type scale, where a higher value corresponds to higher importance.[14] We use respondents' answers to this question as a measure for their work preferences. Thus, our dependent variable is ordinal and ranges from 1 to 7. In the sample, the 95% confidence interval for the 'preference for work' measure ranges from 6.03 to 6.10 across all individuals. The overall mean is 5.84 in the West sample (5.94 for men and 5.74 for women) and 6.39 in the East sample (6.41 for men and 6.38 for women). We consider our "stated preference measure" as a useful intermediary between measures extracted from lab and field experiments and other "revealed preference measures", such as labour force participation. It complements existing studies on the historical, cultural and social determinants of actual female labour force participation (e.g. Alesina et al., 2013; Fernández, 2013; Fogli and Veldkamp, 2011; Fortin, 2005; Fernández et al., 2004) and can help circumvent both external and internal validity issues. Using Spearman correlations, we confirm that our variable is indeed positively associated

[14]Respondents are presented with various "spheres of life" in a random order, and asked to evaluate them independently from each other. We use individual ratings of other "spheres" as controls, see Section 1.4.3 for details. Further descriptive statistics for all preference measures used are provided in Table 1.5 in Appendix 1.8.

with working full time, and negatively with the labour market status 'inactive'. [15] These correlations are significant for both men and women in our sample. They appear to support our interpretation of the individual ratings of the importance of "job and work" as preferences for work, are summarized in Table 1.6 in Appendix 1.8.

While we expect that the political regimes have an effect on gender differences in work preferences, other confounding factors, such as potential differences in socio-demographic characteristics, in individual preferences for alternative means of time use, and in the macroeconomic environment have to be taken into account. Before we introduce our strategy to deal with these confounders in the following subsection, and proceed with our main analysis, we discuss a potential measurement issue in our dependent variable. One concern is that what East Germans respond to the survey question "We would like to know how important each of these spheres of life is for you: Job and work" may not be a truthful representation of their preferences. As described in Section 1.2, avoidance of work was considered a criminal act and could be punished with up to five years in prison in the GDR. Combined with a far-reaching mass surveillance of individuals by the Stasi, the answers to the survey question, whose mean for East respondents is close to 7 out of 7 in 1991 (i.e., shortly after reunification), could rather reflect fearful compliance with GDR ideology. To address this concern, we examine other items in the ALLBUS that should potentially present East German respondents with a similar dilemma. If fearful compliance to GDR ideology was driving the answering behaviour, these items, too, should show a concentration of East German respondents on one, supposedly non-dissident, answering option. Our review of a wide variety of items that fall into this category shows that answering behaviour does exhibit great variation, and leaves us fairly confident that the answers to our

[15] The variable is also significantly correlated with an indicator for cohabitating with a partner for both genders, but the association is negative for women and positive for men, which is in line with economic models of intra-household labour division. Finally, correlations with the variable personal monthly income (positive) and the indicator for living with small children (negative) are significant only for women.

question of interest most likely reflect individual preferences for work at the time, rather than fearful compliance.[16]

1.4.3 Estimated model

To investigate the conditional influences of social structure and institutions in the GDR (East) versus FRG (West) on the gender gap in work preferences, we estimate the following OLS model using the pooled cross-sections from 1991, 1998, and 2012[17]:

$$\begin{aligned} WP_i = &\Theta_{1991} + E_i\Gamma_{1991} + Fem_i\Phi_{1991} + (E_i \times Fem_i)\Pi_{1991} + \\ &\Theta_{1998}Y^{1998} + (Y^{1998} \times E_i)\Gamma_{1998} + (Y^{1998} \times Fem_i)\Phi_{1998} + (Y^{1998} \times E_i \times Fem_i)\Pi_{1998} + \\ &\Theta_{2012}Y^{2012} + (Y^{2012} \times E_i)\Gamma_{2012} + (Y^{2012} \times Fem_i)\Phi_{2012} + (Y^{2012} \times E_i \times Fem_i)\Pi_{2012} + \\ &\boldsymbol{X_i}\Lambda + \epsilon_i \end{aligned} \tag{1.1}$$

WP_i denotes our outcome variable, work preferences, and E_i indicates whether respondent i was living within the borders of the former GDR at the time of the interview. Fem_i indicates a female respondent. The indicators Y^{1998} and Y^{2012} are equal one for respondents interviewed in 1998 and 2012, respectively, and zero otherwise. Thus, the coefficients Θ hold the survey-year fixed effects, including the

[16]The 1991 ALLBUS questionnaire contained a number of potentially sensitive questions, e.g. which party respondents voted for in the last federal election (which took place early December 1990). Note that the interviews were conducted May through July 1991, i.e., more than half a year after German unity had been established and more than a year after the GDR ministry of security had been dissolved. In the 1991 cross-section, only around 6% of respondents reported having voted for the PDS, the political successor of the leading socialist party in the GDR, the SED. Moreover, when asked whether there was anything about the former GDR that they were feeling proud of, 41% of East German respondents stated that there was *nothing*. Other important questions include rehabilitation of former Stasi agents (about 34% agree people should no longer be asked whether they have been working for the Stasi, 60% disagree), how strongly respondents identify with the former GDR and its population (more than 50% report not feeling connected at all or only very weakly connected).

[17]We also estimate all analyses to be presented using a binary probit model, for which we code the dependent variable to take on the value one if the respondent assigned at least a value of six to the importance of work, and zero otherwise. The results are qualitatively unaffected.

constant (Θ_{1991}) and time shifts (Θ_{1998} and Θ_{2012}) in WP_i for the reference group, West German men. The coefficients Γ and Φ, respectively, capture the divergence in the time trend for East German men and West German women with respect to the reference group, i.e., the 'regional gap' in men's preferences for work and the 'gender gap' in the West. Our main interest rests with the coefficients Π, which we obtain from interacting the East dummy with the dummy for female respondents and the survey year, thus revealing *regional disparity in the gender gap* directly after reunification (Π_{1991}) and its change over time (Π_{1998} and Π_{2012}). We will refer to this as the gap in the gap, or GiG. $\boldsymbol{X_i}$ is a vector of individual preference-related, socio-demographic and macro-level control variables, all of which allow us to flexibly control for federal state and time heterogeneity. ϵ_i denotes the individual error term.

Note that taking the double difference (by gender *and* region) rules out the potential problem of different response behaviour due to overall cultural or interpretative differences between East and West respondents, regardless of gender. For our difference-in-difference analysis of the GiG, we only need to rely on the sensible assumption that men and women within the Eastern and Western regions interpret the question in the same way. Thus, any potential general "scaling" differences between East and West Germans, which can bias the comparison between men (or women) across regions, would not affect the GiG. Also note that although our focus lies on female preferences, our empirical identification does not rely on the assumption that work preferences of men have not been affected by socialism.

Controls

Our set of controls can broadly be divided into three categories: preference, socio-demographic and macro-level controls. By including the *preference control variables*, we evaluate preferences for work conditional on the importance respondents assign to competing means of time use. The reason that we are interested to learn about the conditional preference for work is that we assume individuals to

maximize their utility under constraints, i.e., to prioritize according to their preferences when choosing labour supply. We thus include respondents' evaluation of the importance of: family and own children, leisure and relaxation, friends and acquaintances, and relatives.[18] A probit regression using these life aspects as independent variables confirms that a higher *ceteris paribus* evaluation of the importance of work in our sample corresponds with a higher probability of being employed.

The second set of *socio-demographic* controls poses a few challenges. Even though stated preferences have been shown to causally affect labour market outcomes (Fortin, 2008; Humlum et al., 2012; Zhan, 2015), one might be concerned about the potential endogeneity of, e.g., individual human capital investment and labour market participation decisions. Thus we try to reduce the problem of reversed causality, which may arise even in a natural experiment setting, by including only variables in $\boldsymbol{X_i}$ that cannot be influenced by the individual herself. Among the individual-level controls, this leaves us with the respondent's age and the parents' level of schooling, as well as the father's occupational status.[19] We exclude individuals' employment, income, cohabitating status, as well as information on whether they have (small) children, from the analysis for intuitive reasons. As discussed earlier, these variables are correlated with, and may likely be outcomes of individual work preferences.[20] Note, however, that the results we provide in the subsequent sections are relatively insensitive to the inclusion of these variables (except for employment status, naturally).

Our third set, the *macro-level* control variables, captures a wide range of economic and demographic federal state characteristics in order to mitigate the concern that any regional differences we find in the gender gaps regarding work preferences are merely driven by structural differences in respondents' economic

[18] See Table 1.5 in Appendix 1.8 for summary statistics.

[19] Mothers' occupational status was not recorded in the ALLBUS before 2002. In a robustness check, we include this variable using only the 2010/2012 cross-section and verify that this does not affect our results.

[20] See Table 1.6 in Appendix 1.8.

conditions by virtue of living in a certain federal state. Still today, more than 20 years after reunification, the economic development and labour market conditions in the Eastern states lag behind the West. Goldin (1995) shows that the relationship between economic development and female labour force participation is U-shaped, and thus one of our main concerns is to account for heterogeneity in economic development. Consequently, we include federal-state level per-capita GDP, deflated at the state-level consumer price index, and the share of GDP in agriculture and industry. Since we are interested in East-West differences in the gender gaps with regard to preferences, we also include gender-specific unemployment rates[21], a measure of public childcare availability[22], the share of church members, and, among them, the share of Protestants.[23] We obtain all macro-level variables from official register data (Destatis, 2015).

1.5 Results

1.5.1 The evolution of the gaps

Table 1.3 shows the OLS-estimated coefficients of the preference for work in four different model specifications. Model I displays the results for the fully interacted model without further controls. Models II, III and IV successively add the preference, socio-demographic, and macro-level controls. Model V includes a first

[21]Ideally, we would also want to include gender pay gaps at the federal-state level. Unfortunately, for the years prior to 2006, this information is not available in the register data. We checked the robustness of our findings for the 2012 cross-sections, for which we have the administrative information. The results were unaffected.

[22]We constructed this measure from official register data as the ratio of the number of public childcare spaces for children below the age of 7 that have been allocated in a federal state in a given year to the number of children below the age of 7 who then lived in the same state.

[23]Becker and Woessmann (2008) show that, historically, female literacy in Germany first spread in regions with a higher share of Protestant church members. We thus include the share of Protestants as a proxy for different rates of female empowerment, which evolved differentially prior to the German separation, in order to avoid over-estimating the effect of the separation on the 'gap in the gap' in work preferences.

sensitivity analysis which we will refer to in Section 1.6. Shortly after reunification, in 1991, we see that East German residents assign significantly more importance to work than West Germans. A point estimate of 0.61 for the East dummy variable tells us that an average East German man's evaluation of work is more than half a point higher than that of a West German man in the reference year 1991. West German women's preferences for work fall short of West German men's; their average evaluation is 0.288 points lower. Including the preference controls in Model II reduces the East-West difference somewhat, and increases the gender difference, which supports our initial conjecture that there might be scaling differences in response behaviour between East and West, and men and women. It is reassuring to see that the regional gap in the gap (GiG), which is given by the coefficient on the interaction of East and Female, is unaffected by this exercise, as well as by the inclusion of the other control sets.

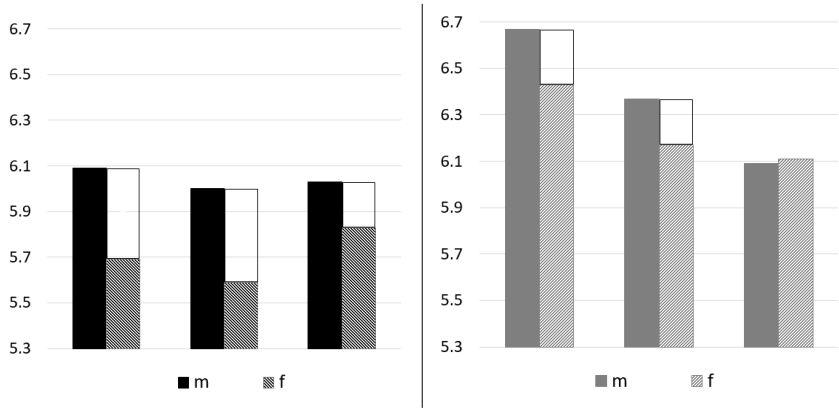
Across all specifications, we see the East-West difference for men shrinking until 1998 (shown by the negative coefficient estimate for $East \times 1998$) and reducing further until 2012 (compare the absolute terms of East and $East \times 2012$). This pattern applies almost equally to women as the $East \times Female$ interaction with both years, 1998 and 2012, remains insignificant. Overall however, women rate work lower than men and the gender gap becomes smaller over time (compare the coefficient for Female and its compensation to almost zero by $Female \times 2012$). This pattern is explicated in Figure 1-1, which displays the predicted means of the work preferences for the two genders in both regions as well as their changes over time, obtained from Model IV, our preferred specification. This figure shows that, mainly due to women's increasing valuation of work, the initial West German gender gap of 0.39 falls to a still significant 0.19 within 20 years. The East German gender gap starts from just above this level (0.24) after reunification and disappears entirely until 2012, due to a stronger downward convergence of men's than women's preferences for work. Figure 1-1 further illustrates that, given these parallel trends in a shrinking gender gap, the GiG remains more or less stable at around 0.15 to

Table 1.3: Preferences for work

VARIABLES	I	II	III	IV	V
East	0.610***	0.568***	0.554***	0.582***	0.510***
	(0.057)	(0.056)	(0.058)	(0.127)	(0.098)
Female	-0.288***	-0.396***	-0.393***	-0.394***	-0.372***
	(0.057)	(0.060)	(0.060)	(0.065)	(0.067)
East x Female	0.156*	0.148**	0.149**	0.150**	0.182***
	(0.075)	(0.068)	(0.066)	(0.067)	(0.057)
1998	-0.121	0.072	0.053	0.077	0.136
	(0.083)	(0.088)	(0.090)	(0.104)	(0.123)
East x 1998	-0.154	-0.170	-0.159	-0.217	-0.143
	(0.135)	(0.136)	(0.126)	(0.129)	(0.102)
Female x 1998	0.019	-0.032	-0.019	-0.018	0.046
	(0.155)	(0.156)	(0.156)	(0.158)	(0.161)
East x Female x 1998	0.046	0.089	0.070	0.068	0.110
	(0.180)	(0.173)	(0.170)	(0.172)	(0.180)
2012	-0.098	-0.075	-0.078	0.010	-0.008
	(0.067)	(0.058)	(0.063)	(0.077)	(0.078)
East x 2012	-0.434***	-0.417***	-0.407***	-0.520***	-0.443***
	(0.085)	(0.074)	(0.073)	(0.088)	(0.095)
Female x 2012	0.226**	0.204**	0.198**	0.200**	0.171**
	(0.083)	(0.084)	(0.080)	(0.084)	(0.087)
East x Female x 2012	0.074	0.067	0.070	0.070	0.028
	(0.119)	(0.119)	(0.108)	(0.119)	(0.108)
Constant	6.020***	2.860***	3.405***	3.773***	4.430***
	(0.047)	(0.272)	(0.324)	(0.394)	(0.500)
Preference controls	NO	YES	YES	YES	YES
Socio-dem. controls	NO	NO	YES	YES	YES
Macro controls	NO	NO	NO	YES	YES
Female LFP 1934	NO	NO	NO	NO	YES
Observations	5,165	5,141	5,141	5,141	5,141
R-squared	0.064	0.213	0.222	0.223	0.224

Note: Robust standard errors in parentheses (clustered at the federal state level). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 1-1: Predicted means for work preferences in West and East by gender and year



	1991	1998	2012	1991	1998	2012
Men	6.09	6.00	6.03	6.67	6.37	**6.09**
Women	5.69	5.59	**5.83**	6.43	6.17	**6.11**
GG	0.39***	0.41***	**0.19*** **	0.24***	0.19**	-0.03
GiG (W-E)	0.15**	0.22	0.22**	0.15**	0.22	0.22**

Note: From left to right, bars show predicted means for West men and women, and for East men and women, based on coefficients from the full estimation model, column IV (see Table 1.3. Unfilled boxes indicate gender gaps that are statistically significant below the 10% level. Calculations of the gender gaps (male – female predicted means) and the GiG (West – East gender gap) are based on coefficients from the full estimation model and may thus entail small rounding errors. Note that the y-axis is zoomed in because more than 95% of our sample assign at least a value of 5 to the importance of work. Stars indicate joint F-test significance at the *** 1%, ** 5%, and * 10% level. Bold figures indicate joint F-test significance (below 10%) for within-group changes over time with respect to the 1991 value.

0.22 scale points over time, indicating that gender-specific preferences in the East and West follow a similar convergence process after reunification, though still at very distinct levels.[24]

The finding of distinct levels in gender differences in preferences for work in the two regions points toward a lasting effect of political systems. Although a gender gap in work preferences is universally present in our sample in 1991, it is significantly smaller among East German respondents. Restricting attention to these 1991-figures, the mere existence of an Eastern gender gap leaves room for two interpretations: One may conclude that, possibly, the GDR policy enforced East German women's participation in the labour force at a higher level than they would have chosen themselves. An alternative explanation may derive from the fact that, despite greater equality of men and women in the East German society, some roles and responsibilities were still divided along gender lines (see Section 1.2). However, we do not observe a convergence of the East gender gap toward the Western level over the course of time after 1991, as we would have expected, had the high participation of women in East Germany been fully diametric to their "true" preferences. Since we find Western levels slowly catching up with the East instead, the second interpretation is arguably more plausible. This is also consistent with the earlier analyses of Maier (1993); Braun et al. (1994) and the "social learning models" proposed by Fogli and Veldkamp (2011) and Fernández (2013). In these overlapping-generations models of social learning, the information signal about the consequences of female participation in the labour force becomes less noisy to the individual female decision-maker as more and more women are observed participating. This may also explain why East German women, with a longer and more comprehensive collective experience of labour market integration compared to West German women, had higher work preferences at the time of reunification.

[24]Note that the change in the GiG is not statistically significant, hence we do not interpret it as an increase.

Having detected differences in the gender gaps in stated work preferences that suggest their malleability by social structure, we would like to validate how these patterns extend to individuals' observed actions (i.e., their employment status). As indicated in Section 1.2, gender gaps in actual labour market attachment, and the GiG in particular, show a very similar pattern in our sample.[25] There is an initial West German gender gap in labour market attachment of 0.29 (women are 29 percentage-points less likely to be active in the labour market, employed or unemployed, than men), which falls to still significant 0.08 in 2012. East Germans, on the contrary, start into reunification with a gender gap in employment of zero. This underlines that East German women, although affected more severely by unemployment than men, kept actively searching, so that the gender gap in labour market *attachment* remained low. It also contradicts the supply side explanation of Hunt (2002) from earlier on while supporting her demand side one, i.e., employer discrimination. Interestingly, we find that in 2012, a gender gap of equal size as in the West has evolved in the East German sample. Due to this development, the initial GiG in employment of 0.29 in 1991 disappears entirely until 2012. One reason for this may be that the preference for work itself is influenced more strongly by cultural legacy, which may still be effective in 2012, while the actual behaviour is shaped more strongly by the then equal institutions and constraints. An alternative explanation would be that gender differences in preferences are driven by older cohorts in the sample, who have had more exposure to GDR institutions, while the impact these differences have on labour market behaviour is most crucial at a younger age, during family formation and child-rearing. Since in 2012 few people of child-bearing age have actually experienced GDR institutions in our sample, differences in labour market behaviour may disappear. To disentangle the mechanisms behind the pattern found in the aggregate analysis, a cohort analysis seems indispensable to gain a better understanding of the "nature of nurture".

[25] An illustration is provided in Figure 1-4 in the Appendix.

1.5.2 How much nurture does it take? A cohort analysis

Having identified an overall nurture effect of the political system of the GDR on gender differences in work preferences, we are now interested in disentangling the sources more precisely: Does the GiG result from pure exposure to different regimes, and can we determine a critical length of exposure for a nurture effect to unfold? And how do work preferences that have been developed in the East German socialist system, but with varying time intensity, continue to evolve under the conditions of the West German market economy?

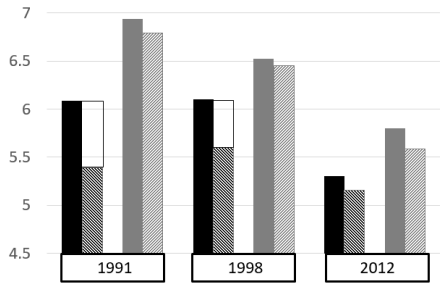
We exploit the exogenous variation in age in our sample to approach these questions since it allows us to compare different subgroups that naturally vary in treatment intensity. East German cohorts differ not only in length of exposure to the GDR system, but also with respect to the point in time of their life when they were hit by the 'shock of reunification' and had to adapt to the dramatic changes that East German institutions and markets were undergoing in the subsequent years. While the youngest respondents in our sample had only experienced their childhood and adolescence in the GDR, and were about to commence their work lives in reunified Germany, the oldest had already been employed for a substantial number of years in the socialist system.

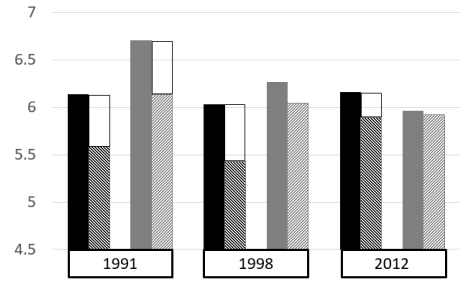
We run separate regressions for four different birth cohorts, defined in a manner to achieve a reasonable degree of variation in their experiences with the GDR regime (cohort effect), and to have enough observations for three pseudo-panel cross sections (life cycle effect). Both dimensions together will allow us to compare the same age groups over time (time effect). For our first group we choose those who were born and raised in the GDR and were impacted by reunification after having spent a substantial share of their working lives in the socialist regime (the eldest cohort 1, born 1953-59). This cohort has consciously experienced both regimes and arguably might have faced the greatest challenges in adapting to the new labour markets and institutions because they were already 32-38 years old around reunification. For the second group we examine those who experienced the

GDR mainly during childhood, adolescence and when entering the labour market but spent most of their working life in reunified Germany with Western labour market institutions (the second eldest cohort 2, born 1960-66, aged 25-31 around reunification). For the third group the transition took place quite early in life (just after adolescence) and adaptation to the Western system may have been less challenging, most likely they did not collect considerable employment experience within the GDR, as they were aged 18-24 around reunification (cohort 3, born 1967-73). Finally, we look at the youngest cohort (cohort 4, born 1974-80), i.e., those without any direct employment experience within the GDR or its labour market, being 16 years of age or younger at the time of reunification, and thus having only experienced the new West German institutions in their working life. Over the full observation period from 1991 to 2012, we are able to observe the older and intermediate cohorts at three points in time – 1991, 1998, and 2012. To achieve this, we include the oldest cohort in 2012 even though at that time they are already older than 50 years, and thus not included in our main analyses. The youngest cohort is only observed twice, as they were too young to be surveyed (and to be active on the labour market) in 1991. We obtain the results presented here using the specification introduced in Section 1.4.3, i.e., the same specification underlying column IV in Table 1.3.

Our cohort regression analyses reveal heterogeneous gender dynamics in preferences for work in East and West Germany across age groups, which result in very different patterns in the GiGs over time (the full set of estimated coefficients is provided in Appendix-Table 1.7). As illustrated in Figure 1-2, among the oldest cohort gender differences in preferences are statistically significant in the West, but not in the East throughout the 1990s (although the GiG is significant only for 1991). Until 2012, when the respondents are aged 53-59, preferences for work decrease markedly for men and women in both regions and, consequently, the GiG vanishes. This is consistent with the psychology literature cited earlier, which shows that individuals' assessment of the importance of work and job aspects de-

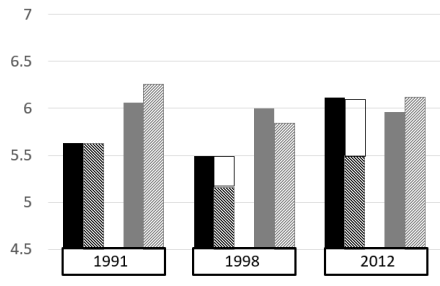
Figure 1-2: Predicted means for work preferences for cohorts 1-4 by region, gender and year



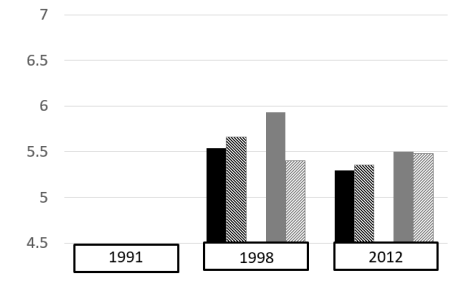


Cohort 1: 1953-59 (N=1164)						
(W|E)	**32-38y.**		**39-45y.**		**53-59y.**	
GG	.70***	.14	.50*	.07	**.15**	.21
GiG	.56***		.43		-.07	

Cohort 2: 1960-66 (N=1409)						
(W|E)	**25-31y.**		**32-38y.**		**46-52y.**	
GG	.56***	.57***	.61***	**.22**	.27**	**.05**
GiG	-.02		.39*		.23	



Cohort 3: 1967-73 (N=1022)						
(W|E)	**18-24y.**		**25-31y.**		**39-45y.**	
GG	.00	-.20	.33*	.15	**.62*****	-.16
GiG	.20		.18		**.78*****	



Cohort 4: 1974-80 (N=492)						
(W|E)	**12-17y.**		**18-24y.**		**32-38y.**	
GG			-.12	.53	-.06	.02
GiG			-.65*		-.07	

■ Wm ▧ Wf ■ Em ▨ Ef

Note: From left to right, bars show predicted means for West men and women, and for East men and women, calculated based on coefficients from the full estimation model (see Appendix-Table 1.7). Unfilled boxes indicate gender gaps that are statistically significant below the 10% level. Calculations of the gender gaps (male – female predicted means) and the GiG (West – East gender gap) are based on coefficients from the full estimation model and may thus entail small rounding errors. Note that the y-axis is zoomed in because more than 95% of our sample assign at least a value of 5 to the importance of work.

teriorates in the middle-age life phase, in anticipation of retirement (Helson and Soto, 2005; Ekerdt et al., 2000; Kalleberg and Loscocco, 1983).

The second and third cohorts of 1960-1973-borns, who experienced reunification in their 20s, head into unified Germany and into their working life without displaying any gap in the gap in 1991, due to similar gender gaps in the East and West, but then develop a statistically significant one over the course of time. Although all groups share a common trend of work becoming increasingly less important over time, by the year 2012, the average GiG in the intermediate cohorts has grown to a similar size as the GiG displayed by the (then same-aged) older cohort 14 years earlier in 1998, and for precisely the same reason: gender preferences converge in the East – predominantly due to the men lowering their valuation – while remaining stable or even increasing in the West. It may seem odd that a GiG for these two cohorts only develops *after* reunification when men and women in both regions live under the same institutions, however, East Germans in both cohorts have significant experience with the positioning of men and women in their native society, which seems to be relevant for their own positioning when they enter child-bearing age.

Interestingly, the youngest cohort displays a different pattern. In 1998, they are in their early twenties and first observed, West German respondents hardly report any gendered preferences, while the East German women appear to care *less* for work than men. Thus, we find a reverse GiG that is remarkable in size. This finding could possibly result from the transition shock, which may have initiated selective migration and/or a reverse social structure effect. With respect to the first effect, it is important to note that this youngest group is the most likely to be highly selected: Hunt (2006) shows that in 1990-2000, among the 18-25-year-olds, East German women were 89% more likely than men to emigrate to the West. Thus, for this cohort, we might expect the gender difference in work preferences in the

West to be biased toward zero, while it could be upward biased in the East.[26] As for the reverse social structure effect, the argument relies on studies showing that the reunification had adverse effects especially on female employment, which may have had dramatic effects on women's position in society. As noted earlier, Hunt (2002) shows that women in East Germany were disproportionately affected by unemployment in the years following reunification, and Witte and Wagner (1995) demonstrate that these women, as opposed to the general trend of sharply declining fertility in the East, showed a *higher* likelihood of having children. This is also confirmed in a recent study by Chevalier and Marie (2017).[27] When we observe the youngest cohort in their thirties in 2012, 14 years later, the East gender gap has reduced to essentially zero. As a result, the reversed GiG has vanished as well.

The different dynamics in the GiG across cohorts also speak to the question which component is more important in shaping preferences – cultural legacy, or own experience with institutions that affect men's and women's positioning in society. The cohort analyses already point to the conclusion that the latter may be more important. To identify the "critical age" or the decisive length of exposure to the GDR for political nurture to have a long-lasting effect, we analyse the GiG in preferences for work in the 2012 cross-section separately for even smaller age groups. As a matter of fact, the GiG as a treatment effect is detectable only for those who are 38 years and older in 2012, i.e., they were at least 16 years of age at the time of reunification. At this age they have cumulated the direct effect of their own exposure to GDR institutions as well as the indirect effect of their parents' exposure through education.

[26]Since in 1998, the ALLBUS did not include the region where respondents lived throughout their adolescence; thus, we rely on residence information at the time of the interview for this cross-section while we will provide selectivity checks for the other years in the next section.

[27]Age at first birth also differed considerably between East and West German women at the end of the 1990s. As a result, the number of children was higher for the early twenties in the East than in the West; female employment rates also differed atypically at that age. A sensitivity analysis including these variables in the regression consequently yields a smaller GiG. Additional support for this family-timing explanation is provided by the subsequent responses of this cohort.

1.6 Sensitivity analyses

So far, our analyses provide evidence that the diverging social structure and cultural context in political regimes has influenced gender differences in preferences decisively. However, they do not necessarily exclude alternative channels through which the effect might be driven. In this section, we explore competing explanations for regional differences in the gender gaps in work preferences, including differences in labour force participation *before* separation, selective migration *after* reunification, and other geographical heterogeneities.

1.6.1 Historical labour force participation rates

Our first objective in this section is to rule out the possibility that historically different working behaviours of women between East and West Germany drive our results. Wyrwich (2017) argues that regional differences in the share of women in the labour force is not significantly higher in 2010 than in 1925 and can therefore not be attributed to the socialist legacy, but is rather a phenomenon of pre-socialist tradition.[28] To address this, we include historical female labour force participation rates at the federal state level into our model. We obtain these data from the statistical yearbook (Statistisches Reichsamt, 1936) and group the 1934 provinces by the modern day federal state boundaries they fall into and calculate average yearly participation rates from the four months (March, June, September, December) reported in the yearbook. The estimation results differ very little by the inclusion of these data, as we can see in columns IV and V of Table 1.3. The estimated

[28] An alternative explanation, which Wyrwich (2017) cannot rule out based on his analyses, is that the socialist regime did increase women's labour force participation relative to men's in the short run, but that this effect has worn off until 2010 (he does not consider points in time that are close to unification). This would actually be in line with the conclusions we draw from our own analyses, namely, that institutions, not cultural legacy influence gender differences in – stated and revealed – preferences. Even though the East effect in preferences for work in our main analysis is still detectable in 2012, we show that it is driven by older cohorts who have experienced GDR institutions. In actual labour market activity, we, too, find that it has already worn off in 2012, but is detectable shortly after reunification in 1991.

GiG becomes slightly larger and statistically more sound in 1991 (compare the coefficient estimates for $East \times Female$ in the two columns). The estimates for $East \times Female \times 1998$ and $East \times Female \times 2012$ show that the GiG hardly changes over time. It remains statistically significant until 2012 – with or without considering historical data.

1.6.2 Youth vs. residence and mover analysis

In this section we investigate possible selection issues resulting from East-West migration. Although we can essentially exclude selective migration before or during separation as a threat to identification in our analytical set-up (see section 1.2), we may have reason to be concerned about the migration flows after reunification. Right after the fall of the wall (and via Czechoslovakia and Hungary even before that), a substantial labour migration from East to West began. The migrants were highly selective in terms of education level, gender, and, presumably, labour market attachment (Hunt, 2006). If highly educated East German women with an over-proportional labour market attachment (for the GDR) comprise a substantial portion of our West German residence sample, our results for the East-West gaps are likely to be underestimated. To examine this, we take advantage of the fact that, for two of our cross-sections (1991 and 2012), ALLBUS respondents provide information on the federal state in which they lived when they were 15 years old.

The results columns 1 and 2 of Table 1.4 serve as the reference for our main analyses presented in Section 1.4; as before, we use the current residence to sort respondents into the East and West categories.[29] In columns 3 and 4, we sort them according to the region they lived in during their adolescence. Finally, for the migrant analysis in columns 5 and 6, we restrict our sample to those respondents who live in the West at the time of the interview, i.e., by construction must have migrated to the West if they report having spent their youth in an Eastern federal

[29]Note that the coefficient estimates differ slightly from those presented in Table 1.3 since, for the sake of comparability, we re-estimate the model excluding the 1998 cross section.

Table 1.4: Youth vs. residency: The gender gap in preferences for work by region and year

	East Residency		**East Adolescence**		**E-W Migrants**	
Work	N=3509		N=3344		N=1927	
	1991	**2012**	**1991**	**2012**	**1991**	**2012**
GG West	0.36***	**0.15*** **	0.38***	**0.12***	0.40***	**0.12****
GG East	0.21***	**-0.09**	0.20***	**-0.05**	-0.62**	**0.44**
GiG (W–E)	0.15*	0.24***	0.18**	0.16	1.02***	**-0.32**

Note: Calculations of the gender gaps (male – female conditional evaluations) and the GiG are based on coefficients from the full estimation model and may thus entail small rounding errors; see Appendix-Table 1.10, Columns I-III.
In the results columns 1 and 2, we replicate the estimation presented in Section 1.5.1, Table 1.3, but excluding the 1998 cross-section for comparability. Respondents who currently live in the West form the reference group, while for those who live in the East the East indicator is one. We apply the same model in columns 3 and 4, but here the East indicator is one only for those respondents who report having lived in the East at the age of 15, and zero for those who report having lived in the West. For columns 5 and 6, we restrict our sample to those respondents who live in the West at the time of the interview, i.e., the reference group consists of people who live in the West and also report having spent their youth there, and the East group consists of people who now live in the West and report having spent their youth in the East.
Stars indicate joint F-test significance at the *** 1%, ** 5%, and *10% level. Bold figures indicate joint F-test significance (below 10%) for within-group changes over time with respect to the 1991 value.

state – and thus examine the gender gaps in preferences between "lifelong" West Germans and East-West migrants. Since the latter have been exposed to the socialist system and its institutions during a rather formative period of their lives, before they decided to migrate (for employment reasons presumably), we expect to observe even greater preference gaps between this group of migrants and the lifelong West sample.

The composite effects in Table 1.4 show the gender gaps in preferences for work by region and year, and reveal the resulting GiGs over time. The results support our previous findings. Comparing the estimations using the adolescence information to the estimations using residence information confirms that the selection bias within the East German population seems partly negligible, as the gender gaps as well as the gaps in the gaps we obtain are very similar. Compared to the original residency sample, the gap in the gap in 2012 is no longer statistically significant and is smaller in magnitude (though still of similar size as the original one in 1991). This is likely a result of elevated migration flows within the country in more recent years, allowing the respective population samples to intermix and individuals to self-select, possibly also with regard to their work and family model preferences, so that the positive Western and negative Eastern gender gap are both smaller.

As the estimation results for East-West migrants show, a selection bias due to the inflow of employment-oriented female migrants from East Germany can be traced among those residing in West Germany in the year 1991. As a result, we see reverse gaps between female and male East migrants (comparing the positive GG East in the column E-W Migrants), with the moving women being much more work-oriented than the men. The resulting 1991-GiG in preferences for work amounts to a full Likert-scale grade (1.02). With a changing composition of migrants, the initial GiG in preferences for work disappears statistically until 2012 (when the point estimate even shows a reversal to -0.32).

1.6.3 Geographical heterogeneities

Another concern we want to address is whether our estimated "East effect" on gender differences in preferences, rather than representing a general East German particularity, applies also to other geographical distinctions or may be due to only one specific Eastern federal state. Despite controlling for the heterogeneous environments with our set of macro control variables, the effect we observe may be driven by some environmental differences the respondents are exposed to in a particular region or GDR state rather than by general exposure to a different political regime. Naturally, the GDR states were not homogeneous in terms of industry structure, economic power, etc. – but neither were the West German states. The GDR regions also varied in distance to the inner-German border and by reception of West German radio and TV channels. Variable exposure to West German programs also implied a natural variation in exposure to respective norms (with regard to the division of work within the household) that may have contributed to preference formation. A study by Hyll and Schneider (2013), for instance, indicates that TV consumption in the former GDR was positively correlated with material aspirations. We approach this issue by applying the same regression analysis as before, but differentiating the GiG for the five former GDR states. If the effect were distributed homogeneously across federal states, we should observe positive and similarly sized effects for each of the Eastern federal states. Overall, our results are largely in line with the GiG we found for the aggregated East population as documented in Appendix-Table 1.8. Particularly in 2012, none of the state GiGs differs significantly from the one observed in 1991.

Another geographical causality exploration is a placebo test on a different division than the East-West divide. By distinguishing between northern and southern German states, we choose a distinction that is not only orthogonal to the East-West dimension, but also almost perfectly coincides with another historical division: the mapping with Prussian districts. Hence, the (never-Prussian) states of Bavaria, Baden Wuerttemberg, Thuringia and Saxony constitute the southern group and

the remaining (predominantly Prussian ancestry) states form the northern reference group. The GiG between these two groups cannot be rejected from being zero in 1991, neither in 2012, which is an indication that the preference divide follows recent political separation (and nurture) rather than more ancient settings (see Appendix-Table 1.9).[30]

1.7 Discussion & conclusion

In summary, we find that women differ systematically from men in their preferences for work on average, but over the whole observation period the difference is much smaller in East Germany than in West Germany. We detect a regional gap in the gender gap (GiG) around the time of reunification that persists in 2012, even though the gender gaps are decreasing in both parts of the country. These findings confirm that political regimes can shape gender differences in preferences substantially. The conclusion that the GiG is most likely caused by exposure to differing political and social systems is supported by our comparison of labour market and family outcomes during and after separation and investigation of potential differences before separation based on historical data sources. The historical analysis delivers support for the identification strategy based on the German separation and reunification not only for the present chapter, but also for other studies relying on this assumption. An important concern was to exclude alternative explanations. For example, differences in the individual preferences for work may be driven by structural differences in the East and West German labour markets today or institutions closely linked to them, such as the availability of public childcare. We show that our findings remain robust to the inclusion of an exhaustive

[30] A refinement of the North-South placebo test leads us to the special case of the Western German state of Bavaria which had similarly large non-marital fertility rates as the Eastern German regions in the 19th century, and still has today (Klüsener and Goldstein, 2014). Our analysis does not support a long-lasting impact of differential non-marital fertility on work preferences of female Bavarians (estimation results upon request). We thus conclude that the two phenomena are likely independent.

set of covariates from official register data that allowed us to flexibly control for time-variant differences at the federal state level. Furthermore, in sensitivity analyses we confirm that the effect is neither driven by differences in female labour force participation before separation nor selective migration after reunification nor geographical specificities. A placebo test supports our conclusion that the preference divide mirrors recently separated institutions in FRG and GDR rather than earlier historical influences, e.g. of the Prussian kingdom.

Having verified that the GiG in stated preferences for work also extends to revealed preferences (i.e., labour supply decisions) in 1991, but no longer in 2012, we employ cohort analyses to gain a better understanding of the mechanism through which short term interventions shape preferences. The analyses reveal that age at exposure and length of exposure are important determinants of the size of the effect, which is not detectable in 2012 among respondents who were younger than 16 years in the reunification year. More specifically, at the time of reunification, a GiG in preferences for work existed only among the 1953–1960 born cohort. Across all cohorts, regional differences in the gender gaps seem to widen over the life cycle, since East German women and men exhibit rather similar preferences for work at any age, whereas the gender gap in West Germany either remains stable at a higher level, or increases around family formation age. For the youngest cohort, born 1974–1980, a GiG did not develop at all. We interpret this pattern as further evidence for the malleability hypothesis, which seems to be working mainly through own experience with institutions that promote gender equality and to a lesser extent through intergenerational transmission of cultural legacy.

Our findings contribute to a better understanding of how gender differences in preferences evolve and align with real (labour) market outcomes observed in East and West Germany.[31] In dealing with both external and internal validity issues, our analysis of stated preference dynamics with regional variation complements the

[31] A cross-country comparison on time trends from areas with similarly distinctive political systems would be a another insightful step, but beyond the scope of this chapter and is thus left to future research.

experimental literature on gender differences in preferences. Using real-world work preferences, it demonstrates the power of preference manipulation in a natural experiment setting. By studying two societies that share a common past and cultural identity up to an exogenously imposed separation and by exploiting variation in treatment intensity (length of exposure to the GDR) that can hardly be replicated in a lab or field experiment, we are able to disentangle distinct mechanisms (experience with GDR institutions vs. GDR legacy) that promote cohort-specific patterns in the differential gender gaps in East and West after reunification.

Our insights into the mechanisms that determine gender differences in preferences, and thereby in economic decision-making, are particularly relevant for the design of equalizing policies. Understanding "the nature of nurture", in particular its potential to influence preferences, may serve as a guide in devising effective strategies to target gender inequalities in labour market outcomes. Our results indicate that relatively short-term changes in social structure can have detectable effects already. The finding that preferences for work vary systematically with the political and institutional setting during one's youth, underlines their particular impact at the height of preference formation. Does it need a change in the role of women as severe as promoted in socialism, or will smaller policy reforms yield similar effects? Although a conclusive answer is left to future research, our results provide evidence that policy measures, like female quotas in supervisory boards or parental leave months dedicated to fathers, that actively change the roles that women and men are assigned to in a society, may have a positive effect on gender equality in labour market outcomes. As suggested by this research, such positive effects may not only work directly through increasing a gender's respective representation, but also indirectly, as they have the potential to shift female and male preferences to fill these roles.

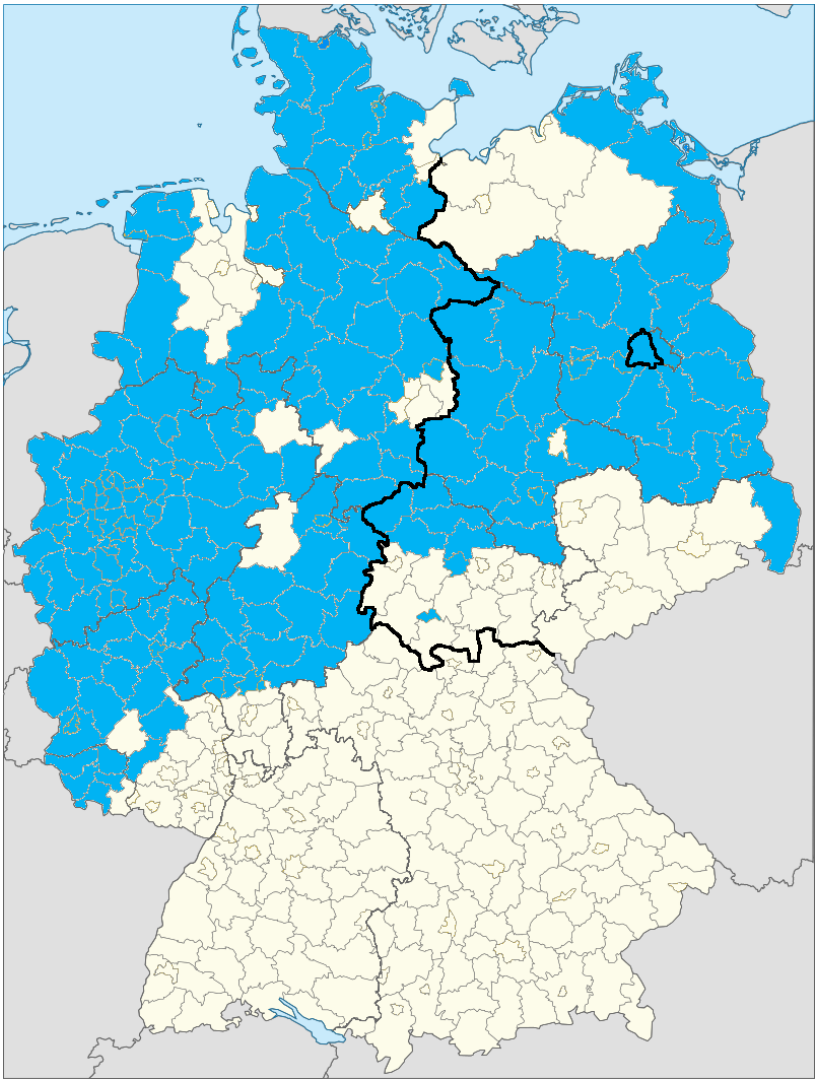
1.8 Appendix

Table 1.5: Descriptive statistics for preference measures for work and other aspects of life

Variable[a]	**n**	μ	σ
Job and work	5165	6.06	1.22
Own family and children	5166	6.28	1.4
Leisure time and relaxation	5167	5.78	1.22
Friends and acquaintances	5165	5.72	1.18
Relatives	5165	5.1	1.52

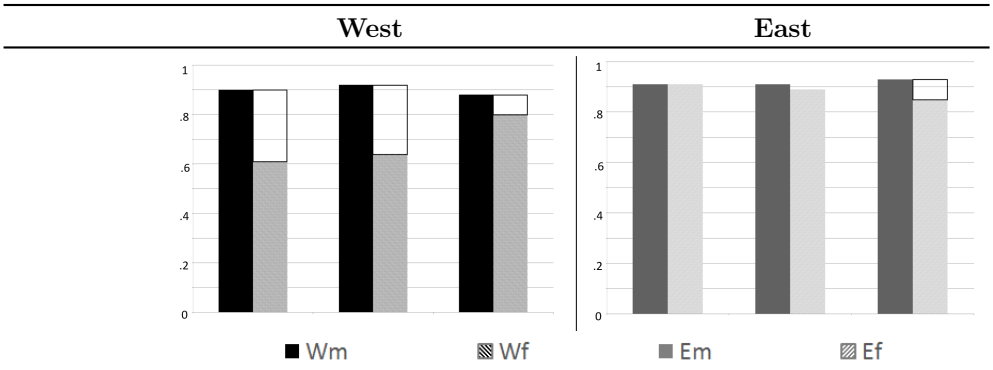
[a]Question: The cards here list various spheres of life. We would like to know how important each of these spheres of life is for you.

Figure 1-3: Geographical overlap of Prussian counties in 1896 within the contemporary German district borders



Own illustration based on Meyers Konversations-Lexikon, 1896: 184.

Figure 1-4: Predicted means of labour market activity in West and East by gender and year



	1991	1998	2012	1991	1998	2012
Men	0.90	0.92	0.88	0.91	0.91	**0.93**
Women	0.61	0.64	**0.80**	0.91	0.89	**0.85**
GG (M-W)	0.29***	0.28***	**0.08*** **	0.00	0.02	0.08***
GiG	0.29***	0.26**	0.00	0.29***	0.26**	0.00

Note: Calculations of predicted means are based on coefficients from the full estimation model, where the dependent variable "active" is 1 if a respondent reports to be either employed or unemployed, available from the authors upon request. Stars indicate joint F-test significance at the *** 1%, ** 5%, and * 10% level. Bold figures indicate joint F-test significance (below 10%) for within-group changes over time with respect to the 1991 value.

Table 1.6: Preferences for work, Spearmen correlation coefficients

	Women	N	**Men**	N
Income[1)]	0.10***	2355	-0.02	2186
Young children[2)]	-0.09***	2671	0.03	2494
Full time work[3)]	0.14***	2653	0.13***	2491
Inactive[4)]	-0.22***	2561	-0.17***	2448
Cohabitating[5)]	-0.06***	2670	0.13***	2489

Note: Stars indicate significance at the *** 1%, ** 5%, and * 10% level.
1) Income: Self-reported personal monthly income, deflated at the state-level consumer price index.
2) Young kids: Indicator for respondents living with at least one child below age 7.
3) Full time employed: Indicator for respondents working full time.
4) Inactive: Indicator for respondents who are neither (self-)employed nor actively looking for a job.
5) Cohabitating: Indicator equal 1 for respondents cohabitating with a partner.

Table 1.7: Cohort Analysis: Gender gap in preferences for work by region and year

VARIABLES	1953-59	1960-66	1967-73	1974-80
East	0.846***	0.567***	0.427*	0.394
	(0.201)	(0.154)	(0.215)	(0.366)
Female	-0.699***	-0.556***	-0.004	0.119
	(0.120)	(0.103)	(0.120)	(0.178)
East x Female	0.558***	-0.018	0.199	-0.649*
	(0.135)	(0.133)	(0.191)	(0.341)
1998	0.210	-0.064	0.243	—
	(0.199)	(0.203)	(0.263)	—
East x 1998	-0.424**	-0.337	0.079	
	(0.194)	(0.217)	(0.217)	—
Female x 1998	0.204	-0.054	-0.325	
	(0.265)	(0.163)	(0.203)	—
East x Female x 1998	-0.132	0.405	-0.022	
	(0.285)	(0.252)	(0.334)	—
2012	-0.834**	0.174	0.520	0.321
	(0.286)	(0.330)	(0.379)	(0.560)
East x 2012	-0.351	-0.763***	-0.584**	-0.194
	(0.294)	(0.168)	(0.245)	(0.315)
Female x 2012	0.551**	0.286	-0.619***	-0.063
	(0.230)	(0.185)	(0.114)	(0.284)
East x Female x 2012	-0.625	0.243	0.583*	0.575
	(0.364)	(0.290)	(0.287)	(0.618)
Constant	3.037	2.222*	2.602***	-0.387
	(2.231)	(1.179)	(0.866)	(1.763)
Preference controls	YES	YES	YES	YES
Socio.-dem. controls	YES	YES	YES	YES
Macro controls	YES	YES	YES	YES
Observations	1,164	1,409	1,022	492
R-squared	0.250	0.260	0.219	0.211

Note: Estimates from the full model, estimated separately for each of the four cohorts. Robust standard errors in parentheses, clustered at the federal state level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.8: GiGs for individual East German states

	1991	**2012**
Full sample	0.177**	0.161
Brandenburg	0.153**	0.324*
East Berlin	0.072	-0.200**
Mecklenburg Pomerania	0.143	0.091
Saxony	0.115	0.143
Saxony-Anhalt	0.256**	0.195*
Thuringia	0.262*	0.200*

Note: Calculations based on the full model using the 1991 and 2012 cross sections (see Appendix-Table 1.10, Column IV). Respondents are sorted into the federal states according to the residence information they report for when they were 15 years old. In addition to the single 'East dummy', the model includes separate interactions with the female dummy for each Eastern federal state, which are displayed.
Robust standard errors in parentheses, clustered at the federal state level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Changes over time are not statistically significant at conventional levels.

Table 1.9: Gender differences in preferences for work in North and South

	1991	**2012**
GG North	0.288***	**0.045**
GG South	0.236***	0.131
GiG (N–S)	-0.052	0.086

Note: Calculations of the gender gaps (male – female predicted means) and the GiG based on coefficients from the full model (see Appendix-Table 1.10, Column V), using the 1991 and 2012 cross-sections.
Respondents who report having lived in the North at the age of 15 form the reference group, while for those who report having lived in one of four southern German states (Bavaria, Baden-Württemberg, Saxony or Thuringia) the South indicator is one.
Stars indicate joint F-test significance at the *** 1%, ** 5%, and *10% level. Bold figures indicate joint F-test significance (below 10%) for within-group changes over time with respect to the 1991 value.

Table 1.10: Sensitivity analyses

VARIABLES	(I) East Residency	(II) East Adolesence	(III) E–W Migrants	(IV) Federal State Level	(V) South Placebo
F	-0.361***	-0.376***	-0.396***	-0.378***	-0.288***
	(0.061)	(0.060)	(0.072)	(0.061)	(0.075)
East	0.479***	0.513***	0.003	0.485***	—
	(0.125)	(0.098)	(0.219)	(0.106)	—
East x F	0.154*	0.177**	1.016***	—	—
	(0.075)	(0.070)	(0.304)	—	—
2012	0.016	-0.000	-0.062	0.008	-0.190
	(0.078)	(0.094)	(0.107)	(0.099)	(0.123)
East x 2012	-0.470***	-0.376***	0.418*	-0.347**	—
	(0.106)	(0.125)	(0.229)	(0.129)	—
F x 2012	0.210**	0.260***	0.274**	0.263***	0.243**
	(0.083)	(0.087)	(0.099)	(0.088)	(0.085)
East x F x 2012	0.087	-0.016	-1.333***	—	—
	(0.121)	(0.132)	(0.294)	—	—
South	—	—	—	—	0.018
	—	—	—	—	(0.072)
South x F	—	—	—	—	0.052
	—	—	—	—	(0.077)
South x 2012	—	—	—	—	-0.035
	—	—	—	—	(0.099)
South x F x 2012	—	—	—	—	-0.138
	—	—	—	—	(0.115)
Brandenburg x F	—	—	—	0.153**	—
	—	—	—	(0.069)	—
East Berlin x F	—	—	—	0.072	—
	—	—	—	(0.141)	—
Sax.-Anhalt x F	—	—	—	0.256**	—
	—	—	—	(0.095)	—
Meck.-Pom. x F	—	—	—	0.143	—
	—	—	—	(0.118)	—
Saxony x F	—	—	—	0.115	—
	—	—	—	(0.070)	—
Thuringia x F	—	—	—	0.262*	—
	—	—	—	(0.140)	—
Brandenburg x F x 2012	—	—	—	0.171	—
	—	—	—	(0.175)	—
East Berlin x F x 2012	—	—	—	-0.272	—
	—	—	—	(0.182)	—
Sax.-Anhalt x F x 2012	—	—	—	-0.061	—
	—	—	—	(0.148)	—
Meck.-Pom. x F x 2012	—	—	—	-0.052	—
	—	—	—	(0.137)	—
Saxony x F x 2012	—	—	—	0.028	—
	—	—	—	(0.122)	—
Thuringia x F x 2012	—	—	—	-0.062	—
	—	—	—	(0.194)	—
Constant	4.824***	4.940***	3.779***	4.873***	5.113***
	(0.347)	(0.275)	(0.880)	(0.278)	(0.455)
Preference controls	YES	YES	YES	YES	YES
Socio.-dem. controls	YES	YES	YES	YES	YES
Macro controls	YES	YES	YES	YES	YES
Observations	3,509	3,344	1,927	3,344	3,509
R-squared	0.167	0.172	0.113	0.173	0.161

Note: Columns report estimates from the full model, using either residency at the time of the interview (I), or reported residency during adolescence for the full sample (II) and for Western residents only (III). Column (IV) reports the estimates from a model including separate interactions of the Female dummy and each of the Eastern federal states. Column (V) shows the estimates for the placebo test, for which we assign respondents from the West German states Bavaria and Baden-Württemberg and the Eastern states Saxony and Thuringia to a placebo treatment group "South" (the remaining states serve as a reference group). In both (IV) and (V) we use reported residency during adolescence to sort respondents into the federal state. All models use the 1991 and 2012 cross sections.
Robust standard errors in parentheses, clustered at the federal state level. Significance levels are indicated by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Chapter 2

Gender matters in language and economic behaviour[1]

Abstract

We study the role of gendered language in shaping economic behaviour theoretically and empirically. A key insight derived from our theoretical model and a systematic literature review regards the methodological task of separating the cognitive effects of language from those of institutions and other cultural influences. In the empirical part of the chapter, we illustrate the severity of the problem by employing the epidemiological approach, a strategy applied in the literature to combat the issue. Using a European data set, we find that behaviour seems hardly related to speaking a gendered language, regardless of the empirical specification, and argue that this finding is driven by self-selection into migration for first generation immigrants and intergenerational transmission of social norms and values for the second generation. Overall, we conclude that the epidemiological approach is not suited to solve the issue of co-evolvement of culture and language, and, moreover, does not seem to solve the important issue of self- and parental selection.

[1]This chapter is joint work with Miriam Beblo and Eva Markowsky. Valuable comments by Gerd Mühlheußer and Arne Pieters are gratefully acknowledged.

2.1 Introduction

Despite vast improvements in gender equality over the past several decades, women are still less active in the labour market and spend fewer hours in paid work than men (UNDP, 2015: 108). Although this is true in most countries of the world, the extent to which women integrate into the labour market varies widely. For instance, the 2013 employment rate of women ranged from only 26% in South Africa to almost 80% in Iceland (OECD, 2016). In the past decade, economic researchers have begun to investigate the role of social norms and culture in shaping these differences (e.g., Akerlof and Kranton, 2000; Duflo, 2012; Fortin, 2005; Fernandez and Fogli, 2009) and, even more recently, the role of language (e.g., Mavisakalyan, 2015; Gay et al., 2017). The present chapter contributes to this endeavour in three respects: i) We provide a conceptual framework for understanding the effects of language on behaviour; ii) We review empirical studies in the economics literature that investigate the link between gendered language and gender gaps in various economic outcomes, and identify challenges for establishing causality; and iii) We present an empirical analysis of European labour markets that confirms the challenges we have outlined in determining causal relationships.

The conceptual framework we developed draws on insights from (psycho-) linguistics, identity economics, and institutional economics. Since linguistic representations of biological sex vary across languages (Stahlberg et al., 2007), they differ in the degree to which they compel speakers to reference gender. These cross-linguistic differences in gender marking could potentially help explain variations in individual economic behaviour through two plausible channels: The first runs from grammatical structure through cognition to speakers' choices. The second plausible channel runs from institutions and culture to individual behaviour. Assuming that culture and institutions (including language) co-evolve, the gender marking of a language is indicative of the degree of gender inequality in a society. While not mutually exclusive, the second channel presents a challenge for empiri-

cists trying to provide evidence for the first: a direct, causal effect of language on behaviour.

Our review of empirical papers in the economics literature confirms that the two channels are not easily discriminated, neither conceptually nor empirically. Even though most authors are interested in the direct, cognitive effect of language on behaviour, the possibility that effects are driven by culture and, in some cases, institutions, cannot be dismissed. The problem is most apparent in cross-country studies that compare economic outcomes at the country level (e.g., the gender pay gap) or at the individual level (e.g., labour force participation), and find correlations between these outcomes and the gender marking in the countries' dominant languages. Seeking a remedy, researchers have turned to the epidemiological approach (Fernandez and Fogli, 2009) and studied the behaviour of immigrants, for whom the institutional context at the time of making decisions of interest no longer coincides with the institutional context in which their language evolved. In the context of gendered language, however, this approach creates new problems of endogeneity that we find warrant further cautious investigation.

In the empirical section of the chapter, we present analyses that confirm the identification challenges we carved out. We apply the epidemiological approach to study the labour market behaviour of immigrants in Europe and show that, despite its advantages in separating the effects of culture from those of institutions, it is not well-suited to tease out the cognitive effects of language from those of culture. To this end, we construct a dataset that is more comprehensive than the ones used in previous studies, and allows us to compare different linguistic indicators that distinguish between gender-intensive and gender-neutral languages. We use several waves of the European Social Survey (ESS) and merge them with linguistic data regarding the respondents' languages. We obtain information about four features of grammatical gender for each language from the dataset of the World Atlas of Language Structures (WALS) (Dryer and Haspelmath, 2013), and complement

this data with information from further linguistic sources for languages missing from the WALS database.

Our systematic analysis of the labour market behaviour of migrants in Europe shows that measuring the impact of language is a challenging undertaking. Conclusions regarding the extent to which language plays a role in shaping behaviour vary, depending on the linguistic indicator used, how the sample is constructed (first- vs. second-generation immigrants), and how language is assigned (most commonly spoken at home vs. dominant language in country of origin). Generally, our results do not confirm the leading hypothesis in the literature, which is that behaviour is more consistent with gender stereotypes for speakers of a gendered language. Instead, our application of the epidemiological approach suggests a weak correlation between behaviour and speaking a gendered language. While for men we find a null effect regardless of whether we look at first or second generation immigrants or whether we assign language by the dominant language in the country of origin or by the language respondents speak at home, women's labour force participation is significantly related to the dominant language for some definitions of gender marking only. If anything, the estimated impact of a gendered language on labour force participation is positive; i.e., female respondents whose parents emigrated from countries with gender marking in their dominant language tend to work more. We argue that these results are consistent with the notion that language reflects cultural traits, and that men and women who select into migration (first generation) are more likely to reject the culture in their country of origin, and may transmit this attitude to their children (second generation). While the epidemiological approach addresses the co-evolution of culture and institutions, it cannot address the interlocking of culture and language, and neither does it seem to overcome self-selection and parental selection issues.

The chapter is organized as follows: Section 2.2 lays out insights from linguistics and psycho-linguistics, from which we derive a conceptual behavioural economics framework for understanding how gendered language can affect behaviour in Sec-

tion 2.3. Section 2.4 presents an overview of economic empirical studies on the subject and highlights inconsistencies in the results, as well as challenges to their interpretation. In Section 2.5, we present our empirical analysis, including data description and empirical approach. The results in Section 2.6 support the conclusions we derived from the theoretical framework and the systematic literature review. Finally, Section 2.7 offers a summary and concluding remarks.

2.2 Theory: The (psycho-)linguistic perspective

Our conceptual framework builds on insights from linguistics and psycho-linguistics, which are described in the following two subsections.

2.2.1 The linguistics of de-constructing grammar

Gender is embedded into the linguistic structure of a language in different ways, as both a grammatical category and a linguistic representation of sex. Generally, linguists distinguish between formal grammatical gender and semantic natural gender (MacKay, 1999). Stahlberg et al. (2007) differentiate among three types of language based on linguistic representation of biological sex: grammatical gender languages, natural gender languages, and genderless languages. In grammatical gender languages—e.g., Spanish, French, and German—sex is coded as a grammatical category. Every noun is assigned either female, male, or (in some languages) neutral gender, and articles, adjectives, pronouns, and in some cases, even verbs must agree with the gender of the noun to which they refer (Boroditsky et al., 2003, , among others). For the majority of personal nouns, grammatical gender corresponds with the sex of the person referred to. Thus, a distinction such as "waiter/waitress," which is the exception in the English language, exists for nearly every personal noun in grammatical gender languages. Consequently, these languages involve frequent reference to gender when verbalizing statements about human beings (Stahlberg et al., 2007). Such references are less frequent in

natural gender languages, because they include no explicit grammatical marking of sex, and gender is used mostly semantically (Konishi, 1993). Examples are English and many Scandinavian languages. With a few exceptions, as in the aforementioned example of "waiter/waitress," person pronouns can refer to both women and men. Only gendered pronouns require referencing sex (he, she) (Stahlberg et al., 2007). Genderless languages—e.g., Finnish and Hungarian—feature grammatical gender neither in their noun system nor in pronouns. Gender is expressed through lexical means only (e.g., in the words for father, mother, sister, brother, and the like) and references to sex are therefore extremely rare (Stahlberg et al., 2007).[2] Some researchers point out that a strict division between these categories is not possible since natural gender languages can feature aspects of grammatical gender languages and vice versa (Konishi, 1993; MacKay, 1999).

In the World Atlas of Language Structures (WALS), the authors Corbett (2013a,b,c) and Siewierska (2013) provide practical guidance for the classification of languages as "gendered" or "neutral" by pointing out four grammatical features to determine the pervasiveness of biological sex in individual languages:

1. Sex-based (SB): Does a language feature a gender system based on the biological distinction between female and male, as opposed to other distinctions (e.g., animacy) (Corbett, 2013b)?

2. Number of genders (NG): How many genders does a language's gender system contain? The French language, for example, features two genders—feminine and masculine—while German has a third gender: 'neuter.' Other languages have more than three or no genders (Corbett, 2013a).

[2] According to Stahlberg et al. (2007), entire language families fall into these three types: (1) Grammatical gender languages: Slavic, Germanic, Romance, Indo-Aryan, and Semitic languages; (2) Natural gender languages: English and the Scandinavian languages; and (3) Genderless languages: Uralic, Turkic, Iranian, Sinitic, and Bantu languages.

3. Gender assignment (GA): How is gender assigned to nouns? Assignment in languages with gender systems can be based on semantic rules only or on both semantic and formal rules (Corbett, 2013c).

4. Gendered pronouns (GP): Are personal pronouns gender-specific? In some languages, personal pronouns can be used with any gender, while others assign different pronouns for each gender category in some or all grammatical persons (Siewierska, 2013).

In the economics literature, the degree of gender marking in languages is predominantly classified using these four grammatical features, as discussed in Section 4. Meanwhile, in the subfield of psycho-linguistics, theories of a causal influence of language on cognition have spurred a rapidly growing empirical research agenda (see Casasanto, 2016, and Haertlé, 2017 for recent overviews). We will touch on these theories in the next section before using our insights to develop a theoretical framework that accounts for language as a determinant in economic decision-making.

2.2.2 The psycho-linguistics of thinking for speaking

The idea that language structures influence the cognitive process dates back to the end of the 19th century, although it gained the most attention through the infamous Sapir-Whorf hypothesis of the 1950s (Casasanto, 2016: 158). According to this hypothesis, also known as Linguistic Relativity or Linguistic Determinism, a person's native language determines her perception, thought, and action (Whorf, 1956). While most researchers have rejected the hypothesis in its strongest, original formulation, a modified version put forward by the psychologist and linguist Dan Slobin (1987, 1996) has found much empirical support. Relative to the original hypothesis, Slobin and other authors suggest a more moderate effect of language on cognition, in the sense that "language differentially favors some thought processes

over others" (Hunt and Agnoli, 1991: 377-378). According to Slobin (1996), language requires "thinking for speaking"; i.e., a cognitive process in order to verbalize experiences. As a result of cross-linguistic differences in grammatical structures, languages vary in the degree to which they draw the speaker's attention to certain aspects. (Slobin, 1996: 71) illustrates this idea by means of the statement "The man is sick." Some languages mandate that the speaker indicates, using grammatical terms, whether the sickness is temporary or chronic, while speakers of other languages would have to specify whether the man was moving or at rest, and so on. He argues that such obligatory reference to certain categories will affect the representation of these categories in the speakers' mind – at least in the process of generating or receiving verbal messages. He refers to gender as another example of semantic features that require obligatory reference in some languages but not in others (Slobin, 2003: 160).

Such a powerful influence on the thought process would not only stem from languages' grammatical idiosyncrasies, forcing speakers to reference certain categories yet ignore others, but from languages' ubiquity in everyday life. Humans are "almost constantly involved in preparing, producing, and interpreting verbal messages" (Slobin, 2003: 158). Moreover, as Casasanto et al. (2004) argue, people might form cognitive habits while "thinking for speaking" which means that these habits become a routine even at times when they are not preparing verbal utterances. Although the authors study mental representation of time, the argument extends to grammatical gender. If a language forces speakers to sort objects into certain categories according to its grammatical rules, it essentially defines which categories deserve the speaker's attention. With respect to grammatical gender languages, this would imply that speakers experience the social categories of female and male as much more relevant than speakers of gender-neutral languages.

2.3 Theory: The economics perspective

Based on the (psycho-)linguistic considerations, we now derive a behavioural economics framework for how gendered language can affect behaviour.

2.3.1 Identity and the salience of gender

If the frequent categorization by gender mandated in grammatical or natural gender languages can make gender as a social category more salient, it can conceivably affect economic decision-making as well. In this section, we build on Akerlof and Kranton's (2000) identity theory to formalize the argument. Mavisakalyan and Weber (2017) propose several channels through which cross-linguistic differences could influence an economic agent's decision-making, many of which we incorporate into our model. First, different languages may convey different amounts of information or influence the cognitive availability of information. Second, languages could influence the individual's preferences. Third, the categories in which agents conceptualize economic decisions may be influenced by the grammatical categories of the agent's language. Finally, they mention salience, i.e., the possibility that different languages may focus the agent's attention on different "actions, outcomes, or states." The authors' economic model integrates social norms as a key determinant of individual behaviour into a standard utility maximization framework. Social norms become powerful because individuals internalize them; i.e., they become an integral aspect of individual preferences. Specifically, the identity utility of an individual—a member of a social category—is directly affected by her ability to match the behavioural prescriptions for that category.[3] The authors illustrate their theory against the background of several examples, with gender inequality in the labour market being the most relevant in the context of the present chapter. Therein, the relevant social categories are "man" and "woman" and the

[3] An agent's identity utility can also be affected by externalities, i.e., by other people's behaviour. This creates additional social pressure to comply with prevalent societal norms. We will discuss this aspect further in the next section.

social norms attached to these categories influence identity utility by prescribing behaviour that is considered appropriate, e.g., in the labour market.

Notably, Akerlof and Kranton (2000) emphasize that "an individual may be mapped into several social categories (e.g., individual j is both a 'woman' and a 'professional')" (718) and that "when an individual's identity is associated with multiple social categories, the 'situation' could determine, for example, which categories are most salient" (731). Following this reasoning and the psycho-linguistic perspective, we hypothesize that speaking a gender-intensive language, which forces the speaker to pay more attention to the social category of gender, increases the salience of this category. Assuming that individuals' (economic) choices are generally influenced by prevalent gender norms and behavioural prescriptions, speakers of gender-intensive languages should be more likely to match behavioural prescriptions for the category of gender and thus more likely to make gender-stereotypical decisions.

We modify Akerlof and Kranton's (2002) schooling model to formalize this notion. While in the AK model the salience of identity is given exogenously, we allow it to depend on language. For simplicity, we follow the original set-up in the AK paper as closely as possible, except that we change the interpretation of the parameters to apply to the context of labour market choices of men and women, rather than schooling choices of "jocks", "nerds", and "burn-outs". Thus, we assume that an individual's utility directly depends on her efforts in the labour market and the monetary returns to it. Denote by e_i the individual's effort in the labour market, measured in hours worked and thus observable to the researcher as labour supply.[4] $c(e_i)$ is the pecuniary costs of effort, including forgone leisure or home production. A standard utility function would thus describe utility as a function of income and effort only: $U_i(we_i, c(e_i))$, where w is the wage rate per

[4]We are aware that effort supply can vary per unit of time, an argument put forward by Becker (1985). However, for the sake of following AK's model set up as closely as possible, we keep the variable effort rather than using time supplied to the market, and assume that effort can be measured empirically as hours worked.

unit of effort. If we assume men and women in a society to be identical, i.e., to differ neither in terms of their returns to effort, effort costs, nor preferences, all agents in the economy will supply the same amount of effort to the labour market. This framework can be augmented by integrating social categories and the corresponding behavioural prescriptions such that the standard utility function described above becomes: $U_i(we_i, c(e_i), I_i)$, where $I_i = I_i(e_i; e^G)$ is the identity component of the utility function. Note that the relevant behavioural prescriptions for specific social categories here are assumed to be the "ideal" effort levels for individual i of gender G (where $G = f, m$), which we denote by e^G. Consistent with the AK model, the standard utility and the identity component are combined using a weighting function. For convenience, we employ the simple functional form used by AK, as it suffices to illustrate the main insight, and express utility for individual i as follows:

$$U_i^G = (1-p)\left[we_i - \frac{1}{2}e_i^2\right] + p\left[-\frac{1}{2}(e_i - e^G)^2\right] \tag{2.1}$$

The parameter p $(0 \leq p \leq 1)$ denotes the weight that is placed on the conformity with one's social category. Individuals experience an identity penalty whenever their effort supply deviates from the socially prescribed, "ideal" effort level for their gender, e^G. The two polar cases of p are worth examining. When $p = 0$, social categories, and thus identity concerns, are irrelevant to the individual's utility. To maximize utility, subjects simply choose the effort level that equalizes the marginal benefit to the marginal cost of expending an extra unit of effort, so $e_i^* = w$. When $p = 1$, the individual will choose an effort level that matches the social prescription exactly, $e_i^* = e^G$. Assuming for now that the behavioural prescriptions for the social categories "man" and "woman" are exogenously given, and that $e^f < e^m$, men would supply more effort to the labour market than women.

In fact, for $e^f < w \leq e^m$, the conclusion carries over for any $p > 0$, as the optimal effort level is given by $e_i^* = (1-p)w + pe^G$.

The linguistic theories on "thinking for speaking" suggest that p may be a function of the decision-maker's language, $p(l)$. We assume that it increases in the frequency with which it forces speakers to reference the category of gender. For given gender norms, i.e. e^f and e^m, that are assumed to be the same across countries for now, this implies that effort choices are more in line with gender norms for speakers of gendered languages than for speakers of gender-neutral languages. To see this, imagine that l were continuous, i.e., language's gender marking may fall anywhere on the spectrum between completely neutral and fully gendered. Then, the first derivative of the optimal effort level, e_i^*, with respect to l can be examined to see how higher gender marking affects effort supply:

$$\frac{\partial e_i^*}{\partial l} = (e^G - w)\frac{\partial p}{\partial l}. \tag{2.2}$$

Because we assumed that $e^f < w \leq e^m$, and that the perceptiveness to gender norms, p, increases in l, this means that the difference in optimal effort supply between genders increases as l increases. In the following subsection, we will consider the inter-relatedness of language and gender norms in a society by adding this dimension to the framework.

2.3.2 Institutions and societal gender norms

The model we introduced above predicts cross-country variation in the gender gaps in labour supply emerging from variation in the grammatical gender feature of languages. One of the simplifying assumptions was that gender norms are invariant across countries. In the presence of heterogeneity in norms, cross-country variation in the gender gap may still result from the mechanism described by the model, i.e., gendered languages increasing the salience of gender categories, so long as the

gender norms in a country are orthogonal to the presence of a gendered language. This assumption, however, appears hard to justify. It seems much more plausible that language and institutions, as well as culture, co-evolve. In the linguistics literature, several studies emphasize that culture is a problematic confounder when attempting to isolate the causal effect of language on cognition (e.g., Beit-Hallahmi et al., 1974; Roberts and Winters, 2013; Beller et al., 2015).[5] Consider low female labour force participation and language gender marking as an example: If a society, historically, strongly relied on a gendered division, this may have shaped both its language and institutions over hundreds of years. The gender marking would then simply reflect the (historical) pervasiveness of gender norms in the speech community, which will also be confounded with current institutional characteristics, such as low support for female labour force participation. Thus, at the country level, the singular impact of language on cognition and individual choices is hard to establish because p and the difference $e^{m} \breve{} e^{f}$ will be positively correlated. Recently, economists were able to provide evidence that this concern is justified: Galor et al. (2017) show that specific geographical conditions, determined well before industrialization, predict certain characteristics of the local languages, such as the existence of a future tense or the presence of grammatical gender. Because these conditions have also been shown to affect institutions and the way society has been organized historically, the authors conclude that language structures reflect past human experience and ancestral cultural traits.[6]

[5]Even languages themselves may be inter-related through cultural evolution. Common ancestors or spillovers can cause neighboring languages to share linguistic features (Roberts et al., 2015)

[6]Notably, some studies in the economics literature go one step further and assume that social norms and language are linked only through historical conditions, dismissing the possibility that language may affect cognition entirely. These studies argue that grammatical features of languages, because they evolve slowly over time, are only correlated with values and social norms, but do not directly affect current institutional outcomes. For example, Tabellini (2008) exploits the presence of a certain grammatical feature, the rules of pronoun use, as an instrument to estimate the causal effect of morality as a social norm on the quality of government. Givati and Troiano (2012) use the number of gender-differentiated pronouns to instrument gender discriminatory attitudes to show the causal effect of the latter on maternity leave policy. Thus,

We can incorporate the idea that language may be shaped by social norms into our model by allowing for the prescribed effort levels, e^G, to depend on language as well. Extending the model in this manner yields[7]:

$$U_i^G = (1 - p(l))\left[we_i - \frac{1}{2}e_i^2\right] + p(l)\left[-\frac{1}{2}(e_i - e^G(l))^2\right] \tag{2.3}$$

While this modification does not alter the condition for an optimal effort level, it *does* change the first derivative of the effort supply function with respect to the gender marking of the language. Formally, it becomes:

$$\frac{\partial e_i^*}{\partial l} = (e^G - w)\frac{\partial p}{\partial l} + p(l)\frac{\partial e^G}{\partial l} \tag{2.4}$$

The first term is the same as before; since we have assumed that (i) the perception of gender identity, p, increases as language gender marking l increases, and (ii) gender norms prescribe an effort level e^G for women that is lower than what is optimal in the absence of gender norms (w), it is negative for women and zero or positive for men. The second term reflects the fact that gender norms, and thus the difference in behavioural prescriptions for men and women, are stronger in societies where language gender marking is higher. Therefore, this term, too, is negative for women and positive for men. Consequently, differences in the gender gap in labour supplies across countries will intensify for two reasons: a stronger gender marking in the local language increases both the gender difference in behavioural prescriptions and the salience of gender as a social category. Empirically,

these authors take an even stronger stance and refute the possibility that language may affect cognition, because this would mean it could, in fact, also have an influence on current policies and institutions.

[7]To keep matters tractable, we use a static model to illustrate our main point. The fact that languages and norms co-evolved in the past affects the individual decision-maker in the present only inasmuch as both social gender norms, e^G, and the weight she places on gender identity, p, are determined by l.

we can no longer say what causes the greater gender gaps in labour supplies, since the effect of language on cognition and the feedback effects between language and social norms are indistinguishable. As will be discussed in more detail in the following section, empiricists have been looking for remedies to mitigate this problem. One potential solution is to study immigrants who live in an institutional context that is different from their home country, and for whom language and institutions therefore do not correlate.

Studying the behaviour of immigrants, however, can introduce new pitfalls to the identification of a causal "cognition" effect of speaking a gendered language on behaviour. To illustrate the issues within the framework of our model, we need to add one important feature: Similar to the AK schooling model, we allow individuals to choose their social category. This may seem unfitting at first, given that the relevant social categories are man and woman, and affiliation with one sex is determined by birth for most people, but we believe this extension will prove useful. While we acknowledge that transgender transitions are also a form of choosing one's gender category, we will not consider such cases here. Rather, we model migration choices as a form of choosing one's gender category: Women (men) may choose whether to live in a society with lower or higher measures of gender equality.

To extend the model in this manner, we need to add one additional component from the AK schooling model. There, identity payoffs are not only determined by students' behaviour, i.e., effort levels, but also by their innate attributes (e.g., looks for the jocks, ability for the nerds). In our application of the model, we include the individual's aptitude in the labour market as an innate attribute. We assume there are two types of individuals who can achieve either high or low returns to effort in the labour market ($w_h > w_l$). We also assume that individuals cannot change the impact their mother tongue has had on their cognition because the salience of gender for their identity, $p_i(l_i)$, is determined during early childhood

and cannot be changed thereafter.[8] Thus, $p_i(l_i)$ is unaffected by the decision to migrate.

We use superscripts to denote high and low gender equality. Thus, l^l (l^h) refers to languages with low (high) gender equality embedded in their grammar, i.e., those that require referring to gender as a category frequently (rarely). Similarly, $e^{(f,l)}$ ($e^{(f,h)}$) refers to the behavioural prescription for women's ideal labour market effort in countries with low (high) gender equality. Since lower gender equality imposes stronger restrictions on women's effort supply, $e^{(f,l)} < e^{(f,h)} \leq w$. For convenience, we assume that women born in countries with a gender-neutral language do not perceive gender identity at all, i.e., $p_i(l_i^h) = 0$. Consequently, these women are indifferent between migrating and staying, regardless of their aptitude. This simplification allows us to focus our attention on women who were born in countries with a gendered language, where, as we discussed earlier, gender norms prescribe higher inequality in men's and women's behaviour. Given these assumptions, and abstracting from moving costs, a woman maximizes her utility by choosing her category (woman in a country with high or low gender equality) and effort level:

$$U_i^h = (1 - p_i(l_i^l))\left[w_i e_i - \frac{1}{2}e_i^2\right] + p_i(l_i^l)\left[-\frac{1}{2}(e_i - e^{f,h}(l^h))^2\right] \tag{2.5}$$

$$U_i^l = (1 - p_i(l_i^l))\left[w_i e_i - \frac{1}{2}e_i^2\right] + p_i(l_i^l)\left[-\frac{1}{2}(e_i - e^{f,l}(l^l))^2\right] \tag{2.6}$$

For certain parameter values of w_l, w_h, $p_i(l_i^h)$, it can be shown that women from countries with lower gender equality stand to gain more from migrating when their returns from supplying effort to the labour market are high. This can be seen by

[8] Psycho-linguistic research suggests that gender marking in language influences children's development of gender identity (e.g., Guiora et al., 1982) and that cognitive effects of grammatical gender are most likely formed at an early age (e.g., Flaherty, 2001).

substituting the optimal effort level, $e_i^* = (1-p)w_\theta + pe^G$, with $\theta = h, l$ into the utility function. The result is the following:

$$U_i(e_i^*) = \frac{1}{2}(1-p)[w_\theta^2 - p(w_\theta - e^G)^2]. \tag{2.7}$$

It is easy to show that utility decreases in the absolute difference between w_θ and e^G. Consequently, when $|w_h - e^{(f,l)}| > |w_h - e^{f,h}|$, and $|w_l - e^{f,l}| < |w_l - e^{f,h}|$, women with high labour market aptitude will prefer to migrate to a country with higher gender equality, while those with lower aptitude prefer staying. This implies that women who migrated from countries with high gender equality are equally likely to be of high or low labour market aptitude, but women from countries with low gender equality might be positively selected in terms of their aptitude. Thus, on average, the latter group of migrant women might even supply more effort to the labour market by either showing a higher propensity to be employed (extensive margin) or working more hours (intensive margin).

The model presented in this section illustrates the various challenges in identifying the causal, cognitive effect of language on behaviour. Against this background, we assess the existing empirical approaches by reviewing the economics literature on the subject in the following section.

2.4 Empirical evidence

The idea that language may influence economic choices has received recent attention by economists. In the most renowned paper, Chen (2013) provides empirical evidence that speakers of "futureless" languages exhibit more future-oriented behaviours (saving money, not smoking, practicing safer sex, maintaining a healthy weight). He argues that they perceive present and future as more closely associated, since their language's grammar does not (consistently) require separating between the two time horizons. His argument can be formalized using a theo-

retical framework similar to the one we presented above, as the mechanisms he proposes go from language through cognition to behaviour. Chen (2013) empirical identification relies on intra-country linguistic variance and shows that, even after controlling for various economic and demographic factors, speaking "futureless" languages is indeed associated with more future-oriented behaviour on all outcomes considered.[9] For the purpose of the present chapter, we will restrict our attention to the evolving literature on the link between gendered language and economic outcomes, and discuss how the evidence it provides allows conclusions to be drawn regarding a direct effect of language on behaviour through cognition. The model we presented in Section 2.3 will serve as a guideline.

2.4.1 Correlation studies

At the country level, several studies find a correlation between the gender marking of a country's dominant language and different outcomes related to gender (in)equality. Mavisakalyan (2015) and Gay et al. (2013) show a negative association between the dominant language's gender marking and women's participation in the labour market. Gay et al. (2013) further document a negative relationship with women's representation in politics as well as their access to credit and land. Santacreu-Vasut et al. (2013) argue that gender marking in the dominant language should be associated with stark gender imbalances among political representatives, and show that these countries are indeed more likely to introduce gender quotas for the lower house of parliament. In a similar vein, gender marking of the dominant language correlates negatively with female participation in

[9]Further evidence for a link between grammatical time reference and future-oriented behaviour is provided by Sutter et al. (2013). In a monetarily incentivized experiment in a bilingual city in Northern Italy, they provide children who speak different languages with an inter-temporal choice. They show that children speaking the "futureless" German language, as used in everyday speech, are much more likely to wait for a larger reward in the future than children who speak Italian, a language that grammatically and practically separates the future from the present. Although the authors cannot separate the effect of language from that of other cultural traits in their experimental design, they provide survey evidence that the effect is not driven by cultural attitudes.

corporate management and with female managers' propensity to lead large teams (Santacreu-Vasut et al., 2014). Using data from a meta-analysis comparing studies of the gender wage gap, van der Velde et al. (2015) show that the gender marking of a country's dominant[10] language is associated with a larger adjusted gender wage gap. Most recently, Hechavarría et al. (2017) find that the gender gap in entrepreneurial activity is higher in countries with gendered dominant languages.

2.4.2 Epidemiological approach

The model discussed in Section 2.3 illustrates why findings of correlations between gender marking in a country's dominant language and measures of gender inequality at the country level do not permit conclusions regarding the causal impact of language on individual behaviour through cognition. Rather, we view them as evidence that gender norms, language, and institutions co-evolve. The authors of these nonetheless intriguing findings mostly acknowledge this limitation explicitly. In search of a superior identification strategy, suited to providing causal estimates of the cognitive effect of language on individual behaviour, researchers have turned to the epidemiological approach introduced by Fernandez and Fogli (2009). Originally, this approach sought to identify the effect of culture on women's behaviour (fertility and hours worked), while muting the effect of institutions. To this end, it focuses on immigrant women who face the same institutions in one host country but come from different cultural backgrounds. Therefore, the key argument is that one can rule out institutional constraints (e.g., the educational system or availability of child care) and the overall economic environment as influential factors. Any remaining differences in immigrant women's behaviour that are systematically related to outcomes of women in their country of ancestry are instead likely caused by culture (Fernández, 2011: 490).

[10]From the paper it is not clear how the authors determine the language of a country. Since they cite various papers with similar approaches, however, we assume that they, too, use the dominant language of a country.

Using a similar line of reasoning, researchers interested in separating the effect of language on behaviour from the effect of institutions and culture, e.g., Santacreu-Vasut et al. (2013), propose using the epidemiological approach to identify the cognition mechanism. To date, most micro-level studies on the relationship between language and economic outcomes follow this recommendation and apply some version of the epidemiological approach (Mavisakalyan and Weber (2017) provide an overview). However, some differences in the methodology used deserve more attention, such as the type of immigrants sampled (first versus second generation) and assignment of grammatical gender marking to the individual (dominant language of country of origin versus language spoken at home). They turn out to play a crucial role in influencing the issues we raise in this section, which we believe cast doubt on the usefulness of the epidemiological approach for the specific problem at hand: separating a causal cognitive effect of language from the effects of institutions *and* culture.

Before we develop our argument, we note an interesting exception among the micro-level studies in this literature. Mavisakalyan (2015) does not study immigrants, but exploits within-country variation in languages spoken, an approach similar to Chen (2013). Using the World Values Survey, she categorizes the language respondents report to speak most frequently at home as highly, mildly, or not gendered. Her main finding is that women speaking a highly gendered language are less likely to participate in the labour market. Moreover, conditional on participation, these women are more likely to be employed part-time. Investigating within-country variation allows her to convincingly separate the effect of language from that of institutions. Notably though, she does not interpret her estimates as causal effects of language on behaviour, as even within the same country, she cannot rule out that gender marking in language is correlated with other unobserved characteristics, such as culture or social norms, which could be the true driver of these differences.

Gay et al., in their 2013 working paper, provide results similar to Mavisakalyan (2015), based on the same data and using the same strategy, but with different indicators for gender marking of the language spoken at home. They show that women speaking highly gendered languages are less likely to be employed and also less likely to work in the agricultural sector. Interestingly, when applying the epidemiological approach to a sample of female immigrants in the U.S., however, they find that, conversely, female immigrants spend more hours in formal employment when reporting that they speak a gendered language at home.[11] The authors argue that the reversal of the effect is due to self-selection into migration, suggesting that the observed women consciously escaped the gender norms of their country of origin. This is consistent with the mechanism we describe in our model: women from countries of origin with low gender equality (which coincide with grammatical gender marking in the languages, as the above cited cross-country studies show) migrate at a higher rate when they have a higher aptitude for labour market participation, while women from countries with more gender equality are equally likely to migrate, regardless of their aptitude.

In a more recent paper, the same set of authors, Gay et al. (2017), extract a large sample of female immigrants in the U.S. from the American Community Survey (ACS), to which they apply further restriction criteria: the women must (a) be married with a spouse present in the household; and (b) report speaking a language other than English. With these restrictions in place, Gay et al. (2017) present different results than those of their 2013 working paper: Among these women, speaking a gendered language is associated with a significantly lower propensity to participate in the labour force. Because the results are robust when controlling for linguistic families, even with the inclusion of origin country dummies, the authors claim to identify the cognitive effect of language on behaviour

[11] Going back to their World Values Survey sample and restricting it to individuals living in countries where the dominant language is gender-neutral, Gay et al. (2013) arrive at a similar conclusion: Women who report speaking a gendered language at home are more likely to be active in the labour market.

through variation in spoken languages among immigrants from the same country of origin. It is not clear, however, that the epidemiological approach is suited to achieve this. It was designed to separate the effect of culture from institutions, and was considered superior to comparing individual behaviour across countries while merely controlling for variables meant to capture differences in institutions. The approach loses this advantage when used to study the causal effect of language on behaviour. If language not only affects cognition but also serves as an indicator for deeply rooted social gender norms, then comparing migrants from different cultural backgrounds, albeit within the same institutional setting and incorporating proxies for culture, suffers from the same methodological problem as comparing respondents across countries while including proxies for institutions. Consequently, the results presented by Gay et al. (2017) cannot be interpreted as any "more causal" than those of Mavisakalyan (2015). Similar concerns apply to the study of Hicks et al. (2015), who find that immigrant women in the U.S. from countries with gender marking in the dominant language allocate significantly more time to household chores, while males from these countries report less time allocated to such tasks.

2.4.3 Dominant versus spoken language

Only at first glance does it seem that using the language individuals actually speak at home, rather than the dominant language of the country of origin, may alleviate the concern of assigning culturally ingrained gender norms. Upon deeper reflection, this approach might cause even greater trouble. For example, considering only immigrants who do not speak the host country's dominant language at home (Gay et al., 2017) raises concerns regarding selection: Speaking the host country's language is an indicator for cultural assimilation, which may in turn be related to gender inequality in the country of origin. If women from countries with low gender equality tend to have higher labour market aptitude on average, they might be more eager to integrate into the host country culture. The language individuals

speak at home is a matter of choice, and very likely related to other choices, e.g., who to marry. The language spoken at the time of the survey may not be the same language spoken during an individual's critical age period, in which effects on cognition are said to manifest (e.g., Flaherty, 2001; Sera et al., 1994, 2002). Imposing the restriction that a spouse must be present in the household further compounds the problem, as the spouse might act as an enforcer of cultural gender norms picked up by the language.

2.4.4 First versus second generation

Finally, Galor et al. (2017) show that female second-generation immigrants in the U.S. are less likely to have attended college when they speak a language with a sex-based grammatical gender system. In the epidemiological literature, studying second-generation immigrants is usually considered superior to studying first-generation immigrants because economic migrants are, on average, more able, aspiring, or in some other way more likely to succeed in the labour market than their otherwise similar counterparts who decided to remain in their home country (Chiswick, 1999). These concerns do not typically apply to second-generation immigrants. However, since "epidemiological studies" of second-generation immigrants have shown that cultural beliefs and values are at least partly transmitted from one generation to the next (e.g., Fernández et al., 2004; Farré and Vella, 2013), these individuals are not necessarily better suited to study the causal effect of language on behaviour. If gender marking of language is an indicator for sexist cultural norms, then by assigning the dominant language of the country of origin to second-generation immigrants, we can essentially measure the "intention-to-treat" effect with ancestral culture. Assigning the language most spoken, on the other hand, measures the effect of self-selection into a certain cultural environment within the host country. Additionally, we face the problem of potentially very differently selected first-generation mothers to second-generation daughters. Mothers originating from countries with low gender equality might be more strongly

selected in terms of their labour market orientation, or their rejection of the gender norms in their country of origin (Abramitzky et al., 2014). As a result, it seems unlikely that using the epidemiological approach to study the behaviour of second-generation immigrants can bring us closer to estimating the causal effect of language.

2.4.5 Language indicators

Yet another issue complicates the comparison of these empirical findings: the variety of indicators used to measure a language's grammatical gender marking. All studies discussed here rely on the four grammatical features relating to gender from the World Atlas of Language Structures (WALS) described above. Most authors (except Mavisakalyan, 2015, see below) dichotomize these features by defining languages as gendered according to the following rules and justifications:

- Sex-based (SB): The grammatical gender distinction is based on biological sex.
- Number of genders (NG): The language features exactly two genders, since the presence of additional gender categories requires less frequent reference to maleness and femaleness (e.g., Hicks et al., 2015).
- Gender assignment (GA): Gender is assigned according to both semantic and formal rules; this makes gender more pervasive and therefore more visible (Hicks et al., 2015; Mavisakalyan, 2015).
- Gendered pronouns (GP): The language has gender-specific pronouns in the third person, as well as the first or second person. If a language's gendered pronouns are only in the third person, the literature treats it as gender-neutral.

Some researchers take these four binary indicators and combine them into a "gender intensity measure" by summing up three or four indicators and, in

some instances, conditioning the sum on the language being sex-based (SB = 1). Appendix-Table 2.5 provides an overview of the studies discussed here with regard to the data used, the sample restrictions, the way languages are assigned to individuals, and the different measures of gendered language. The notable exception is Mavisakalyan (2015), who also relies on WALS data, but uses only the information on gendered pronouns to construct her own indicator. She classifies languages as 'highly gendered' when they feature gender distinctions in third person and first- or second-person singular pronouns, 'mildly gendered' when pronouns are in the third-person singular only, and 'gender-neutral' when neither applies.

In summary, our review of the empirical literature raises several issues that limit the comparability of results across studies. It also raises doubts regarding the suitability of the epidemiological approach to study the causal effect of gendered language on economic behaviour.

2.5 Data and empirical design

In our empirical analysis, we illustrate these unresolved issues with separating the cognitive effect of language from that of cultural factors by considering the impact of the three critical aspects just discussed when assessing the epidemiological approach: (i) the generation of migrants to be studied; (ii) the choice of the language indicator; and (iii) the assignment of languages to individuals.

We exploit the variation in the behaviour of immigrants in Europe who, within one European host country, face the same institutional framework but descend from different cultural backgrounds from various countries of origin. The approach is well suited for removing institutional constraints as confounding factors when investigating the effect of language gender marking on labour market behaviour, and allows us to highlight the issues raised in the preceding chapters related to the intertwining of culture and cognition. We analyse weekly working hours, thus covering both the extensive and intensive margin of labour market participation,

because this outcome has been studied prominently with the epidemiological approach to estimate the causal impact of culture (Fernandez and Fogli, 2009).

2.5.1 Data sources and sample selection

For our empirical investigation, we pool seven cross-sections of the European Social Survey (ESS).[12] The ESS has been conducted bi-annually over a period of 13 years (2002 to 2014) and contains information about migrants in 30 (mostly) European countries, though some waves include respondents in Russia, Turkey, and Israel (ESS, 2018). Our measure of working hours is the self-reported number of hours respondents 'normally work a week (in [their] main job), including any paid or unpaid overtime.' We exclude outliers who report working more than 67 hours weekly, but our results are robust to including these observations.

As outlined in the previous section, most studies on language gender marking and individual behaviour focus on first-generation immigrants. Based on the insights derived from our theoretical framework, we argue that, if language gender marking is an indicator for culturally ingrained gender norms, the selection of migrants might vary with the variable of interest. Furthermore, Fernández (2011) (497) argues for studying second-generation immigrants because, unlike their parents, they are less likely constricted by various obstacles to economic engagement, such as language barriers in their host country. However, as discussed in Section 4, in striving to identify the causal effect of language on behaviour, studying second-generation immigrants raises several new pitfalls. First, the most important conclusions from epidemiological studies of second-generation immigrants relate to the notion that cultural beliefs and values are at least partly transmitted from one generation to the next (e.g., Fernández et al., 2004; Farré and Vella, 2013).[13] If

[12] In the first wave, respondents' spoken languages were not recorded.

[13] Using U.S. data, Fernández et al. (2004) show that the employment status of mothers has an effect on the probability of their sons to be married to working women. Farré and Vella's (2013) work utilizes U.S. data from the 1970s. They find that mothers' attitudes towards working women affects their children's view in this regard as well and that mothers with less traditional

the parental generation, depending on their language gender marking, is composed of different selections in terms of their internalized approval of the social gender norms in their country of origin, the particular norms they are passing on to their offspring are unclear. Second, assigning language to the second generation is even more difficult, as both the dominant language in the country of origin and the language most frequently spoken at home are strongly correlated with culture and cultural assimilation.

In the context of the above considerations, we present a detailed comparison of the behaviours of first- and second-generation immigrants, aiming to highlight the relevance of the issues we raised with studying the link between language gender marking and labour market behaviour, i.e., self-selection into migration and the intergenerational transmission of culture. We therefore split the data set to separately study first-generation (born in a country different from the one where the interview took place) and second-generation (at least one parent born in a foreign country) immigrants. Furthermore, only respondents considered to be of working age (25 to 55 years) are included.

2.5.2 Linguistic data

We follow the literature by using the World Atlas of Language Structures (WALS) Online Dryer and Haspelmath (2013) to categorize languages by grammatical gender marking. In order to attain results that are as comprehensive as possible, we present multiple specifications of our regressions, one for each of the four indicators obtained from the WALS that are most commonly used in the literature as previously described: sex-based (SB), number of genders (NG), gender assignment (GA), and gendered pronouns (GP). A fifth specification features the indicator for the number of gendered pronouns used by Mavisakalyan (2015) (GPM). We also present a sixth indicator (GL), which we consider best suited to detect the

attitudes towards women's role in the labour market are more likely to have daughters who are employed.

linguistic incidence of gender marking in a language based on the following considerations: Three of the above four grammatical features—namely, GP, NG, and GA—are only indicative of gender marking in concurrence with grammatical gender categories being based on biological sex (SB), the fourth feature. They do not indicate gender marking if based on some other distinction, such as animate vs. inanimate nouns. Thus, for our GL indicator, we assign gender marking if the language: 1) has a grammatical gender system that is based on biological sex; and 2) is coded as gendered in at least one of the remaining three indicators.

2.5.3 Language assignment

Another important question is how to assign languages to individuals in the data set. The two strategies used in the literature use either the dominant language of the country of origin or the self-reported language respondents speak (see last column of Appendix-Table 2.5 for an overview). Assigning the dominant language is clearly problematic in the case of multilingual countries with no unambiguously dominant language. Consequently, many authors choose the language respondents report to speak most often in daily life. Assigning the self-reported language, however, might contaminate the analysis with endogeneity. First, language use is a choice; it might depend on whether or not a person lives alone, cohabitates with a partner, or resides with family members or compatriots. Thus, we cannot determine whether the language spoken most often is the dominant language in her own country of origin or that of her spouse or another individual living in the residence. The household constellation, in turn, may be affected by both her culturally ingrained gender norms as well as her labour market behaviour. Moreover, speaking the dominant language of the host country or another language is clearly an indicator for the degree of cultural assimilation in the host country, which in turn is an important determinant of labour market outcomes (Chiswick and Miller, 2015). To further explore this issue of possible endogeneity introduced by the spoken language, we present all of our specifications assigning gender marking at both

the level of the dominant language of the country of origin and the language most often spoken at home. We compile information on countries' dominant languages from The World Factbook (Central Intelligence Agency (CIA), 2017), the Atlas of the World's Languages (Asher and Moseley, 1994), and the Ethnologue (Gordon, 2005). Appendix-Table 2.6 provides an overview on dominant languages in the respondents' countries of origin.

2.5.4 Final dataset and estimated model

Merging all this linguistic information to the ESS data results in a dataset of 7,399 first- and 5,947 second-generation immigrants.[14] Appendix-Table 2.7 presents summary statistics for the two different datasets.

With this data, we estimate the following model:

$$Hours_{icklt} = \beta_0 + \beta_1 GL_l + \beta_2 \boldsymbol{Y}_k + \beta_3 \boldsymbol{X}_i + \beta_4 \boldsymbol{\delta}_c + \beta_5 \boldsymbol{\theta}_t + \epsilon_{icklt} \tag{2.8}$$

$Hours_{icklt}$ represents the weekly working hours of individual i living in host country c, descending from ancestry k, with dominant language/speaking language l, who is observed at time period t. GL is a binary variable, taking on the value of 1 when the respondent's language features gender marking, and 0 otherwise. $\boldsymbol{Y}$ is a vector of characteristics of the respondent's country of origin used in the literature to capture variation in cultural attitudes and norms: The rate of female labour force participation (FLFP), taken from the International Labour Organization (ILO, 2018), total fertility rates (TFR) from the World Bank (2018), and GDP per capita from the United Nations Statistics Division (UN, 2018). $\boldsymbol{X}$ represents a

[14] Because the number of second-generation immigrants in the ESS is rather limited, we also include immigrants of the so-called "1.5th generation," i.e., those that arrived in the host country together with their parents before the age of 14. For those individuals, it seems safe to assume that the decision to migrate and the selection of the host country was their parents' decision and thus exogenous to the respondents, as in the case of second-generation immigrants.

vector of controls at the individual level: age, age squared, educational level of the respondent's mother (one dummy variable for lower than secondary education and one for tertiary education, leaving secondary education as the reference group), and religion (dummy variables for Christian and Muslim faith, as well as Eastern religions, leaving respondents not reporting any affiliation with a religious group as the reference group). $\boldsymbol{\delta}$ is a set of dummy variables for the host countries, which we include to account for heterogeneous institutions and economic conditions in locations across European destinations. Finally, $\boldsymbol{\theta}$ represents a set of time dummies for the survey waves. As the working hours of women and men are estimated within the same model, a female dummy and all interactions of the female dummy with the country of origin characteristics and host country dummies are also included in the estimation (but not all are displayed, results available upon request).

2.6 Results

The estimated coefficients for the model are displayed in Tables 2.2 and 2.1 for first-generation, and Tables 2.4 and 2.3 for second-generation immigrants. One general finding is that the coefficient estimate for the female dummy is always negative and large in size, indicating that, among migrants with a language background not marked by gender, women work fewer hours than men, though this finding is only statistically significant among second-generation immigrants, as the standard errors are too large in the first-generation sample. A second general finding is that, compared to the baseline category of respondents whose mother acquired secondary education, lower maternal education reduces respondents' working hours significantly. Mothers' education is particularly influential in terms of the labour force activity of first-generation immigrants, for whom the coefficient estimate is twice as large as for second-generation immigrants. Similarly, the finding that respondents of Muslim faith work significantly fewer hours than all other religious (or atheist) groups is more pronounced among the first generation than the second.

FLFP in the ancestry country is negatively related to men's current working hours in both generations. A positive effect on women's labour supply is only observed for first-generation immigrants and may be interpreted as a selection effect. The two samples are differentially affected by the ancestry country's TFR. Whereas higher TFR decreases men's and women's working hours among first-generation immigrants, a significantly and substantially negative impact persists only for women in the second generation. The relationship between GDP and working hours is U-shaped for second-generation immigrants. By and large, the estimation results confirm the expectation of a more diverse sample of first-generation immigrants, whose working hours overall seem more responsive to the circumstances in their origin countries than the subsequent generation's. Nonetheless, the labour market behaviour of the second generation still clearly relates to their parents' educational, religious, and economic origins.

Finally, and most importantly for our research question, the estimation analysis reveals that none of the language indicators for either spoken or dominant language is systematically related to the working hours of men. The interaction between language gender marking and being female, however, shows a statistically significant coefficient for some of the indicators. To facilitate interpretation, we plot the marginal effects for women and men, respectively, in Figure 2-1. The top row refers to first, the bottom row to second generation immigrants. Within each row, panels 1 and 2 show the effects on working hours obtained by assigning the spoken or dominant language. The figures show that language gender marking in the spoken or dominant language is hardly related to the working hours of women either. If any, as for the GA indicator in the bottom left panel of Figure 2-1, the effect on hours is positive, meaning that a woman whose dominant language in her parents' country of origin assigns nouns to genders on both semantic and formal rules is observed to work, on average, 1.5 more hours per week than a comparable woman from an ancestry country with a language that assigns gender only on semantic grounds or not at all. The GL indicator has a coefficient estimate of similar

size, meaning that a woman's working hours increase by about the same number when originating from a gendered dominant language country, though the effect is only borderline significant. A general observation is that none of the estimated indicator coefficients supports a negative impact of speaking a gendered language on women's labour market activity. These findings are in line with the implications derived from our model in Section 2.3: if country-of-origin language reflects the local cultural values and gender norms, women who stand to gain from rejecting these norms will select into migration at a higher rate than others. Focusing on second-generation immigrants does not seem to solve this problem, as the relevant cultural attitudes may still be transmitted from parents to their children. Another general observation is that these coefficients, though not statistically significantly different from zero on average, have slightly larger variance among first-generation immigrants than second-generation immigrants, whose working hours are likely more homogenous after having experienced a longer integration period.

Overall, our results confirm the pitfalls of trying to identify causal effects of language on behaviour with the epidemiological approach laid out in the conceptual framework: Because culture and language are intertwined, studying immigrants from heterogeneous language backgrounds does not help separate the causal impact of language. Additionally, in this context, the epidemiological approach seems to suffer from (parental) selection issues in both generations of immigrants.

Figure 2-1: Estimated marginal effects of gendered language on working hours

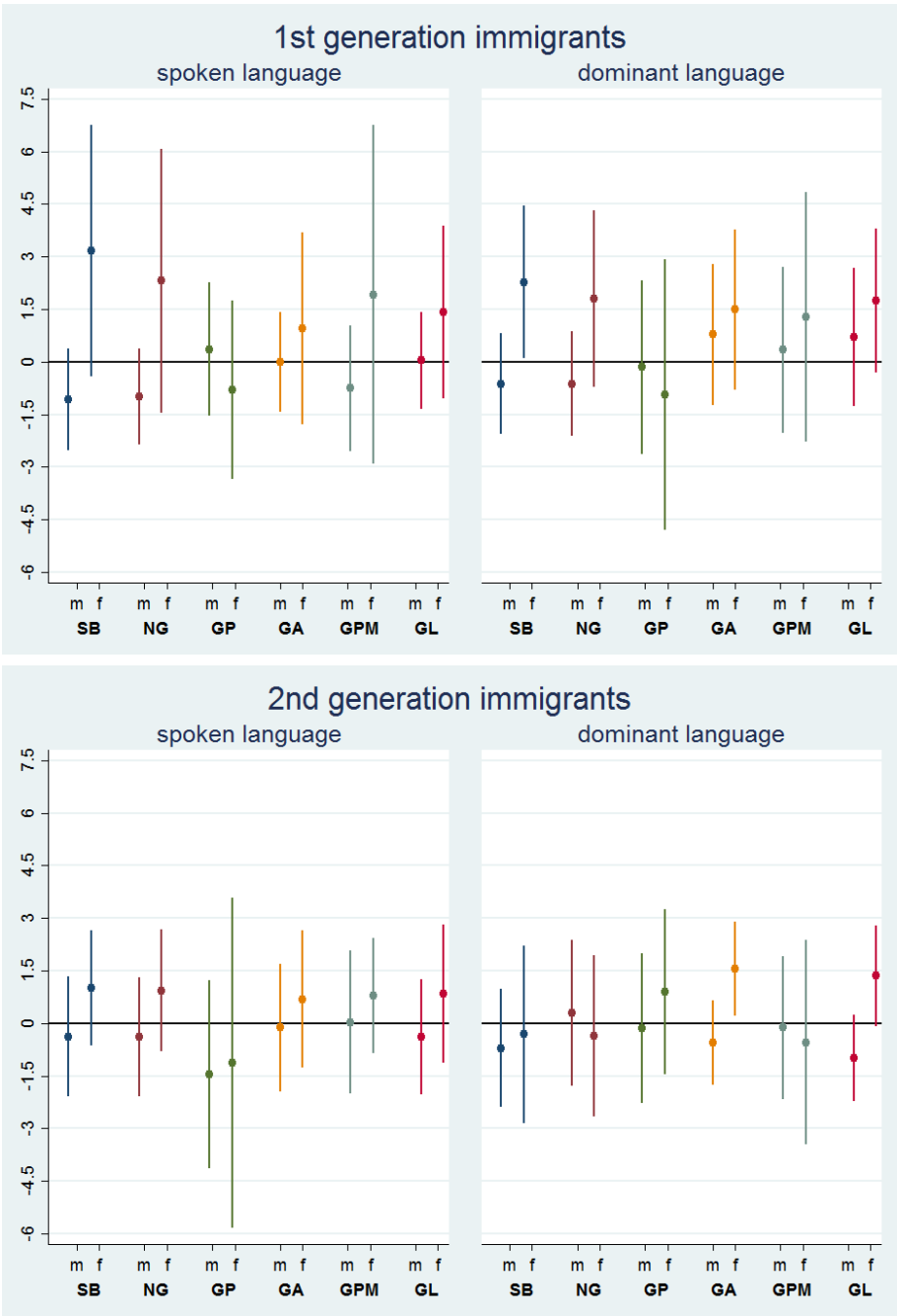


Table 2.1: Regression results for first generation immigrants, spoken language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.186	-13.101	-11.181	-13.079	-12.079	-13.332
	(11.967)	(12.276)	(12.027)	(12.374)	(12.324)	(12.107)
Age	0.358	0.357	0.362*	0.365*	0.361*	0.368*
	(0.216)	(0.216)	(0.216)	(0.215)	(0.216)	(0.215)
Age squared	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother no education[a]	-1.857***	-1.871***	-1.872***	-1.873***	-1.875***	-1.870***
	(0.404)	(0.407)	(0.409)	(0.407)	(0.411)	(0.405)
Mother tertiary education	1.691***	1.697***	1.703***	1.709***	1.706***	1.708***
	(0.355)	(0.354)	(0.354)	(0.354)	(0.353)	(0.356)
Christian	-0.253	-0.237	-0.229	-0.255	-0.227	-0.271
	(0.314)	(0.313)	(0.307)	(0.317)	(0.310)	(0.319)
Eastern	-0.188	-0.239	-0.348	-0.245	-0.317	-0.200
	(0.846)	(0.858)	(0.853)	(0.857)	(0.862)	(0.851)
Islamic	-4.068***	-4.157***	-4.218***	-4.094***	-4.200***	-4.021***
	(0.746)	(0.713)	(0.613)	(0.687)	(0.640)	(0.717)
			Characteristics of the ancestry country			
FLFP	-7.843***	-7.922***	-7.556**	-7.674**	-7.986**	-7.569**
	(2.949)	(2.950)	(3.064)	(2.972)	(3.088)	(2.928)
FLFP x female	10.606**	10.568**	9.797**	10.939**	10.695**	11.311**
	(4.178)	(4.214)	(4.418)	(4.484)	(4.522)	(4.453)
TFR	-0.733**	-0.723**	-0.697*	-0.702*	-0.698*	-0.703**
	(0.355)	(0.357)	(0.364)	(0.356)	(0.362)	(0.347)
TFR x female	0.068	0.055	0.023	0.161	0.006	0.222
	(0.599)	(0.603)	(0.625)	(0.655)	(0.627)	(0.642)
GDP	-0.200	-0.206	-0.210	-0.215	-0.221	-0.212
	(0.198)	(0.198)	(0.198)	(0.181)	(0.199)	(0.184)
GDP x female	0.065	0.088	0.103	0.184	0.123	0.186
	(0.193)	(0.196)	(0.196)	(0.203)	(0.194)	(0.194)
GDP squared	0.006	0.006	0.006	0.007	0.007	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
GDP squared x female	0.002	0.001	-0.000	-0.003	-0.001	-0.003
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
			Language indicator (LI[b])			
LI	-1.081	-0.997	0.361	0.001	-0.291	0.040
	(0.725)	(0.688)	(0.954)	(0.717)	(0.663)	(0.692)
LI x female	3.173*	2.313	-0.793	0.955	0.511	1.429
	(1.799)	(1.888)	(1.273)	(1.373)	(1.502)	(1.234)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.963***	42.903***	41.641***	41.674***	42.169***	41.517***
	(5.396)	(5.396)	(5.282)	(5.297)	(5.313)	(5.290)
Observations	7,399	7,399	7,399	7,399	7,399	7,399
Adjusted R-squared	0.150	0.150	0.150	0.150	0.150	0.150

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Includes mothers with primary education.

[b] From left to right, $LI = SB, NG, GP, GA, GPM, GL$.

[c] All interactions with $Female$ are also included.

Table 2.2: Regression results for first generation immigrants, dominant language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.903	-13.850	-10.656	-15.659	-12.154	-15.818
	(12.107)	(12.273)	(11.867)	(12.111)	(12.015)	(11.927)
Age	0.362*	0.361*	0.363*	0.383*	0.361*	0.389*
	(0.217)	(0.216)	(0.216)	(0.214)	(0.215)	(0.215)
Age squared	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother no education[a]	-1.829***	-1.853***	-1.888***	-1.819***	-1.863***	-1.790***
	(0.414)	(0.415)	(0.417)	(0.398)	(0.417)	(0.401)
Mother tertiary education	1.679***	1.685***	1.717***	1.674***	1.698***	1.667***
	(0.353)	(0.354)	(0.354)	(0.358)	(0.354)	(0.358)
Christian	-0.254	-0.247	-0.230	-0.265	-0.232	-0.278
	(0.311)	(0.310)	(0.310)	(0.311)	(0.307)	(0.313)
Eastern	-0.040	-0.118	-0.431	0.085	-0.219	0.040
	(0.830)	(0.844)	(0.884)	(0.839)	(0.889)	(0.814)
Islamic	-4.073***	-4.139***	-4.219***	-3.793***	-4.172***	-3.728***
	(0.732)	(0.699)	(0.577)	(0.765)	(0.650)	(0.789)
			Characteristics of the ancestry country			
FLFP	-8.145**	-8.237**	-8.023**	-6.047*	-7.739**	-5.896*
	(3.231)	(3.242)	(3.831)	(3.378)	(3.544)	(3.497)
FLFP x female	11.624**	11.494**	8.864*	12.535**	10.796**	13.518***
	(4.528)	(4.551)	(5.295)	(4.866)	(5.047)	(4.973)
TFR	-0.792**	-0.787*	-0.679*	-0.486	-0.695*	-0.511
	(0.393)	(0.395)	(0.373)	(0.427)	(0.366)	(0.424)
TFR x female	0.261	0.217	0.060	0.481	0.013	0.578
	(0.680)	(0.672)	(0.636)	(0.702)	(0.635)	(0.705)
GDP	-0.207	-0.214	-0.222	-0.096	-0.214	-0.137
	(0.197)	(0.198)	(0.206)	(0.179)	(0.206)	(0.164)
GDP x female	0.057	0.087	0.088	0.285	0.122	0.255
	(0.210)	(0.208)	(0.194)	(0.220)	(0.197)	(0.195)
GDP squared	0.006	0.007	0.007	0.003	0.007	0.004
	(0.007)	(0.007)	(0.008)	(0.006)	(0.008)	(0.006)
GDP squared x female	0.002	0.001	0.000	-0.006	-0.001	-0.004
	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
			Language indicator (LI[b])			
LI	-0.624	-0.622	-0.154	0.777	0.008	0.717
	(0.725)	(0.750)	(1.244)	(1.005)	(0.691)	(0.988)
LI x female	2.279**	1.805	-0.940	1.490	0.278	1.755*
	(1.087)	(1.258)	(1.938)	(1.145)	(1.175)	(1.029)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.679***	42.737***	41.851***	38.905***	41.732***	38.937***
	(5.660)	(5.649)	(5.401)	(5.816)	(5.552)	(5.812)
Observations	7,399	7,399	7,399	7,399	7,399	7,399
Adjusted R-squared	0.150	0.150	0.150	0.151	0.150	0.151

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a]Includes mothers with primary education.

[b]From left to right, $LI = SB, NG, GP, GA, GPM, GL$.

[c]All interactions with $Female$ are also included.

Table 2.3: Regression results for second generation immigrants, spoken language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.671**	-13.670**	-13.464**	-13.732**	-13.700**	-13.755**
	(5.463)	(5.463)	(5.446)	(5.464)	(5.451)	(5.455)
Age	0.290*	0.291*	0.301*	0.289*	0.294*	0.291*
	(0.170)	(0.170)	(0.170)	(0.169)	(0.172)	(0.170)
Age squared	-0.003	-0.003	-0.003*	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother no education[a]	-0.870**	-0.871**	-0.877**	-0.871**	-0.882**	-0.874**
	(0.415)	(0.415)	(0.406)	(0.416)	(0.418)	(0.415)
Mother tertiary education	0.183	0.183	0.143	0.182	0.176	0.183
	(0.466)	(0.466)	(0.463)	(0.467)	(0.465)	(0.466)
Christian	0.119	0.123	0.131	0.117	0.151	0.128
	(0.327)	(0.326)	(0.330)	(0.334)	(0.324)	(0.335)
Eastern	0.514	0.509	0.447	0.503	0.459	0.493
	-1.632	-1.631	-1.645	-1.635	-1.612	-1.637
Islamic	-2.921***	-2.926***	-2.835***	-2.897***	-2.965***	-2.934***
	(1.096)	(1.096)	(1.065)	(1.089)	(1.068)	(1.086)
			Characteristics of the ancestry country			
FLFP	-4.928*	-4.931*	-5.172**	-4.908*	-5.070**	-4.975*
	(2.509)	(2.510)	(2.526)	(2.500)	(2.494)	(2.506)
FLFP x female	4.043	4.039	3.733	4.107	4.072	4.140
	(3.600)	(3.597)	(3.492)	(3.579)	(3.527)	(3.575)
TFR	-0.007	-0.007	-0.004	-0.003	-0.009	-0.011
	(0.303)	(0.303)	(0.302)	(0.306)	(0.304)	(0.306)
TFR x female	-1.330**	-1.331**	-1.363**	-1.325*	-1.341**	-1.320*
	(0.657)	(0.658)	(0.653)	(0.666)	(0.655)	(0.665)
GDP	-0.173	-0.173	-0.190*	-0.174	-0.183	-0.175
	(0.110)	(0.110)	(0.113)	(0.112)	(0.113)	(0.111)
GDP x female	-0.466*	-0.465*	-0.468*	-0.456*	-0.457*	-0.459*
	(0.246)	(0.246)	(0.243)	(0.248)	(0.246)	(0.247)
GDP squared	0.006*	0.006*	0.007*	0.006*	0.006*	0.006*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
GDP squared x female	0.013	0.013	0.013	0.012	0.012	0.012
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
			Language indicator (LI[b])			
LI	-0.376	-0.387	-1.456	-0.117	-0.501	-0.394
	(0.859)	(0.852)	(1.339)	(0.912)	(0.745)	(0.821)
LI x female	1.015	0.933	-1.118	0.685	0.316	0.842
	(0.823)	(0.868)	(2.358)	(0.977)	(1.109)	(0.990)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.175***	42.180***	41.852***	41.944***	42.336***	42.212***
	(3.408)	(3.407)	(3.396)	(3.446)	(3.383)	(3.402)
Observations	5,947	5,947	5,947	5,947	5,947	5,947
Adjusted R-squared	0.124	0.124	0.124	0.124	0.124	0.124

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Includes mothers with primary education.

[b] From left to right, $LI = SB, NG, GP, GA, GPM, GL$.

[c] All interactions with $Female$ are also included.

Table 2.4: Regression results for second generation immigrants, dominant language

	SB	NG	GP	GA	GPM	GL
			Individual characteristics			
Female	-13.294**	-13.079**	-14.180**	-17.109***	-14.160**	-16.803***
	(5.488)	(5.305)	(5.537)	(5.489)	(5.459)	(5.514)
Age	0.292*	0.291*	0.291*	0.291*	0.290*	0.295*
	(0.167)	(0.168)	(0.168)	(0.168)	(0.168)	(0.168)
Age squared	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mother no education[a]	-0.908**	-0.868**	-0.867**	-0.873**	-0.869**	-0.887**
	(0.406)	(0.407)	(0.413)	(0.414)	(0.410)	(0.412)
Mother tertiary education	0.201	0.176	0.175	0.195	0.176	0.191
	(0.475)	(0.470)	(0.467)	(0.466)	(0.470)	(0.467)
Christian	0.156	0.131	0.128	0.117	0.127	0.139
	(0.326)	(0.330)	(0.332)	(0.334)	(0.331)	(0.329)
Eastern	0.455	0.469	0.523	0.533	0.524	0.533
	(1.686)	(1.650)	(1.677)	(1.632)	(1.669)	(1.650)
Islamic	-3.193**	-2.883**	-2.900***	-2.857**	-2.895**	-3.040**
	(1.259)	(1.173)	(1.060)	(1.148)	(1.121)	(1.199)
			Characteristics of the ancestry country			
FLFP	-5.481**	-4.667*	-5.086*	-5.576**	-5.007*	-6.573**
	(2.692)	(2.604)	(3.030)	(2.770)	(2.730)	(2.917)
FLFP x female	3.827	3.697	5.375	6.162*	4.520	6.293*
	(3.687)	(3.687)	(3.793)	(3.632)	(3.594)	(3.726)
TFR	-0.048	0.033	0.015	-0.120	0.004	-0.227
	(0.308)	(0.287)	(0.319)	(0.309)	(0.305)	(0.310)
TFR x female	-1.372**	-1.390**	-1.434**	-0.984	-1.357**	-1.016
	(0.646)	(0.644)	(0.644)	(0.669)	(0.662)	(0.643)
GDP	-0.148	-0.183	-0.178	-0.210*	-0.174	-0.212*
	(0.127)	(0.127)	(0.110)	(0.111)	(0.108)	(0.111)
GDP x female	-0.452*	-0.453*	-0.426	-0.333	-0.456*	-0.404*
	(0.252)	(0.248)	(0.267)	(0.245)	(0.249)	(0.237)
GDP squared	0.005	0.007	0.006*	0.007**	0.006*	0.007*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
GDP squared x female	0.012	0.012	0.011	0.008	0.012	0.011
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)
			Language indicator (LI[b])			
LI	-0.713	0.289	-0.133	-0.543	-0.069	-0.979
	(0.843)	(1.044)	(1.071)	(0.601)	(0.540)	(0.617)
LI x female	-0.317	-0.358	0.906	1.545**	0.305	1.352*
	(1.264)	(1.154)	(1.177)	(0.670)	(0.449)	(0.724)
Host country dummies[c]	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Constant	42.719***	41.389***	41.840***	42.919***	41.887***	43.957***
	(3.628)	(3.654)	(3.546)	(3.562)	(3.604)	(3.566)
Observations	5,947	5,947	5,947	5,947	5,947	5,947
Adjusted R-squared	0.124	0.124	0.124	0.124	0.124	0.124

Note: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a]Includes mothers with primary education.
[b]From left to right, $LI = SB, NG, GP, GA, GPM, GL$.
[c]All interactions with $Female$ are also included.

2.7 Conclusion

Our empirical finding of negligible effects of the language indicators are fully in line with the arguments we derived from our theoretical framework and from a systematic review of the empirical literature. Based on (psycho-)linguistic and identity economics considerations, the basic version of our model shows that the extent to which people conform with their socially prescribed roles depends on the gender marking of their language, i.e., the emphasis it places on the category of gender. The extended version of the model incorporates the interrelatedness of language and culture, accounting for feedback effects of increased gender marking on the gender differences in behavioural prescriptions. As a result, culture, norms, and language effects cannot be easily discriminated empirically.

While the widely used epidemiological approach addresses the co-evolvement of culture and institutions, it cannot resolve the interlocking of culture and language. Moreover, in the present context, it does not seem suited to overcome the issue of self-selection and parental selection, which we carved out in a further extension of our model, in which women may choose a country of residence with high or low gender equality and their (socially desired) work effort level simultaneously. One important insight from this model was that gender marking may affect individual behaviour through two channels: (a) grammatical structure by way of cognition; and/or (b) institutions and cultural norms. While not mutually exclusive, these two channels present a challenge for empiricists trying to measure the effect of language on behaviour and establish a causal link.

In our assessment of the empirical literature, we illustrated that the two channels are indeed not easily discriminated. As a consequence, the gender marking of an individual's language does not seem to have a systematic and robust impact on labour market behaviour. None of the language indicators in our application of the epidemiological approach, neither for spoken nor for dominant language, is systematically related to the working hours of immigrant men; this is true for both first- and second-generation immigrant men. Second-generation immigrant

women's working hours are related to only one aspect of gender marking of the dominant language: whether nouns are assigned to genders on both semantic and formal rules. Consequently, our results do not confirm the leading hypothesis in the literature, which is that behaviour is more consistent with gender stereotypes for speakers of a gendered language. Instead, behaviour seems hardly related to speaking a gendered language and, if anything, women work more hours. We argue that these results support the idea that language reflects cultural traits, and that men and women who select into migration as first-generation immigrants are more likely to reject the culture in their country of origin, and may transmit this attitude to their children (second generation). With this chapter, we hope to have contributed to a systematic assessment by shedding both light and doubt on the growing literature aiming to investigate a causal link between gendered language and gender gaps in various economic outcomes.

2.8 Appendix

Table 2.6: Sample countries, languages and language indicators

Country of origin	Dominant language	N	SB	NG	GA	GP	GPM	GL
Albania	Albanian	231	1	1	1	0	1	1
Algeria	Arabic	247	1	1	1	1	2	1
Argentina	Spanish	116	1	1	1	1	2	1
Australia	English	79	1	1	0	0	1	0
Austria	German	147	1	1	1	0	1	1
Belarus	Russian	224	1	1	1	0	1	1
Belgium	Dutch	159	1	1	1	0	1	1
Bolivia	Spanish	31	1	1	1	1	2	1
Brazil	Portuguese	202	1	1	1	0	1	1
Bulgaria	Bulgarian	141	1	1	1	0	1	1
Cambodia	Central Khmer	7	0	0	0	0	0	0
Canada	English	85	1	1	0	0	1	0
Chile	Spanish	50	1	1	1	1	2	1
Colombia	Spanish	72	1	1	1	1	2	1
Costa Rica	Spanish	5	1	1	1	1	2	1
Croatia	Croatian	246	1	1	1	0	1	1
Cuba	Spanish	29	1	1	1	1	2	1
Cyprus	Greek	16	1	1	1	0	1	1
Czech Republic	Czech	202	1	1	1	0	1	1
Denmark	Danish	96	0	1	1	0	1	0
Dominican Republic	Spanish	22	1	1	1	1	2	1
Ecuador	Spanish	86	1	1	1	1	2	1
Egypt	Arabic	67	1	1	1	1	2	1
El Salvador	Spanish	3	1	1	1	1	2	1
Equatorial Guinea	Spanish	2	1	1	1	1	2	1
Estonia	Estonian	61	0	0	0	0	0	0
Finland	Finnish	213	0	0	0	0	0	0
France	French	592	1	1	1	0	1	1
Germany	German	980	1	0	1	0	1	1
Greece	Greek	104	1	1	1	0	1	1
Guatemala	Spanish	7	1	1	1	1	2	1
Honduras	Spanish	6	1	1	1	1	2	1
India	Hindi	193	1	1	1	0	0	1
Iraq	Arabic	192	1	1	1	1	2	1
Israel	Hebrew	15	1	1	1	1	2	1
Italy	Italian	624	1	1	1	0	1	1
Kazakhstan	Russian	206	1	1	1	0	1	1
Kuwait	Arabic	7	1	1	1	1	2	1
Latvia	Latvian	79	1	1	1	0	1	1
Madagascar	Malagasy	13	0	0	0	0	0	0
Mauritania	Arabic	2	1	1	1	1	2	1
Mexico	Spanish	25	1	1	1	1	2	1
Morocco	Arabic	779	1	1	1	1	2	1
Netherlands	Dutch	186	1	1	1	0	1	1
New Zealand	English	19	1	1	0	0	1	0
Nicaragua	Spanish	3	1	1	1	1	2	1
Niger	Hausa	4	1	1	1	1	2	1
Nigeria	Yoruba	77	0	0	0	0	0	0
Panama	Spanish	3	1	1	1	1	2	1
Peru	Spanish	65	1	1	1	1	2	1
Philippines	English	84	1	1	0	0	1	0
Poland	Polish	814	1	1	1	0	1	1

Portugal	Portuguese	453	1	1	1	0	1	1
Qatar	Arabic	1	1	1	1	1	2	1
Romania	Romanian	428	1	1	1	0	1	1
Russia	Russian	1946	1	1	1	0	1	1
Sao Tome and Principe	Portuguese	7	1	1	1	0	1	1
Saudi Arabia	Arabic	4	1	1	1	1	2	1
Slovakia	Slovak	246	1	0	1	0	1	1
Slovenia	Slovenian	39	1	1	1	1	1	1
Suriname	Dutch	92	1	1	1	0	1	1
Sweden	Swedish	223	0	0	1	0	1	0
Syrian Arab Republic	Arabic	85	1	1	1	1	2	1
Thailand	Thai	47	0	0	0	0	0	0
Tunisia	Arabic	164	1	1	1	1	2	1
Turkey	Turkish	590	0	0	0	0	0	0
United Arab Emirates	Arabic	1	1	1	1	1	2	1
United Kingdom	English	779	1	1	0	0	1	0
United States	English	274	1	1	0	0	1	0
Viet Nam	Vietnamese	49	0	0	0	0	0	0
		$\sum$ 13346						

Table 2.5: Literature overview

	Data	Sample	Language indicator(s)	Language assigned
Gay et al. (2013)	World Values Survey (1994-2007)	Native women	All four indicators separately	Dominant language in country of origin;
	New Immigrant Survey (NIS)	1st gen immigrant women, 18-65	NG + SB + GA	Language spoken at home
			NG + SB + GP	
Gay et al. (2017)	American Community Survey (2005-2015)	Married 1st gen immigrant women, 25-49, speaking language other than English at home	SB x (GP + GA + NG)	Language spoken at home
Galor et al. (2017)	US Census (2000-)	2nd gen immigrant women, older than 24, living with their parents	SB	Language spoken by respondent
	American Community Survey (2000-)			
Hechavarría et al. (2017)	Global Entrepreneurship Monitor (2001-2005)	Country-level	All four indicators separately	Dominant language
Hicks et al. (2015)	American Time Use Survey (2003-2012)	1st gen immigrants	All four indicators separately	Dominant language in country of origin
	Current Population Survey (2003-2012)		SB + GP + GA + NG	
			GA conditional on SB	
Mavisakalyan (2015)	World Values Survey (1994-2007)	Native women, 18-65, excluding retired and students	GP, three categories	Dominant language in country of origin;
				Language spoken at home
(Santacreu-Vasut et al., 2013)	Quotaproject.org	Country-level	SB + GP + GA + NG	Dominant language
(Santacreu-Vasut et al., 2014)	World Values Survey (1994-2007)	Country-level	All four indicators separately	Dominant language
	MIX market data		NG + SB + GA	
			NG + SB + GP	
(van der Velde et al., 2015)	Meta-analysis, studies on the gender wage gap	Country-level	Not specified, presumably SB, GA, GP	Not clear, presumably as in Santacreu-Vasut et al. (2013)

Table 2.7: Sample characteristics

	First generation		Second generation	
	mean	s.d.	mean	s.d.
Individual characteristics				
Weekly working hours	37.1	14.41	38.4	12.75
Female (0/1)	0.6	0.50	0.6	0.50
Age in years	40.6	8.31	39.9	8.80
Mother no/primary education (0/1)	0.4	0.49	0.3	0.46
Mother tertiary education (0/1)	0.2	0.38	0.2	0.38
Christian religion (0/1)	0.5	0.50	0.3	0.48
Eastern religion (0/1)	0.0	0.12	0.0	0.09
Islamic religion (0/1)	0.1	0.29	0.1	0.25
Characteristics of the origin country				
FLFP (0-1)	0.4	0.12	0.4	0.14
TFR (number of children born)	2.3	1.05	2.3	1.14
GDP (in 1000 US Dollar)	9.1	9.18	9.6	9.51
Observations	7,399		5,947	

Chapter 3

The power of love: A subtle driving force for unegalitarian labour division?[1]

Abstract

I study couples' specialisation decisions experimentally and examine the gender-specific patterns in labour division arising within heterosexual couples. Eighty participants—20 real couples and 20 pairs of strangers—play a two-stage game, paired up either with their partner or a stranger. I find that women are significantly more likely to give up their income autonomy and perform an unpaid task when playing with their partner rather than with an unfamiliar man. Men's behaviour is not affected by familiarity with their female partner.

[1]Valuable comments by Miriam Beblo, Denis Beninger, Bart Golsteyn, Shoshana Grossbard, Daniel Hamermesh, Martin Heidenreich, Thomas Hills, Ulf Kadritzke, Peter Kuhn, Andreas Lange, Marcus Nöth, Annemarie Paul, Arne Pieters, Helmut Rainer, Thomas Siedler, participants of UHH workshops on experimental economics and of the EALE 2014 meeting, as well as from two anonymous referees are gratefully acknowledged. I am indebted to the University of Warwick for financial support for this project. This chapter is pubished in *Review of the Economics of the Household*, 2015, volume 13(1), pages 163-192.

3.1 Introduction

"Often there are fundamental inequalities in gender relations within the family or the household. (. . .) It is quite common in many societies to take for granted that men will naturally work outside the home, whereas women could do so if and only if they could combine such work with various inescapable and unequally shared household duties. This is sometimes called 'division of labour', though women could be forgiven for seeing it as an 'accumulation of labour'. The reach of this inequality includes not only unequal relations within the family, but also derivative inequalities in employment and recognition in the outside world."
Sen (2001)

The motivation for this study arises from a puzzling observation closely connected to Sen's statement that, after a decade, does not appear to have lost its validity: In most European households, couples do not practice an equal sharing of paid, labour market work and unpaid, household-related work. Instead, despite their improving educational achievements and professional qualifications, women are frequently observed to devote their labour primarily to family-work. Many of them still only become active in the labour market to the extent their remaining capacities allow them to. This is one of the main reasons why even modern developed societies have failed to achieve gender equality in the labour market. Many inequalities persist and hinder female economic independence (Fagan and Norman, 2013; Esping-Andersen, 2009).

In this chapter, I present experimental evidence on couples' decisions on dividing paid and unpaid labour and compare their behaviour to mixed-sex pairs of strangers. The main questions addressed here are whether couples divide labour more often in order to reach efficiency gains when this requires the individual disadvantage of one of the partners, and, if so, whether male and female partners are equally likely to undertake the disfavoured role in absence of individual productivity differences.

Most of the specific gender inequalities observable in European labour markets are interdependent with household-related work which is still predominantly provided by women (Saraceno, 2011; OECD, 2012). In 2012, though with remarkable differences between countries, the average rate of female labour market participation was 62.3% within the EU-27, compared to 74.6% for men. As the household-related workload increases, especially when entering parenthood, this employment gap usually widens: figures for adults aged 25 to 49 provided by Eurostat (2013a) show, when entering parenthood, women's participation in the labour market decreases by about 10 percentage-points while men's increases by the same amount. While the share of male inactives or part-time-employed males in the same age-group, who state "child-rearing" or other family-related duties as the main reason for not seeking (full-time) employment is well below 10% in the EU-27, among females, this share amounts to about 40% (Eurostat, 2013b,c). Consequently, women often face additional career-penalties, such as lower wages, fewer chances for promotion, etc. (European Commission, 2015a). Thus, as a result of gender-specific labour division, we observe women to give up their income autonomy more frequently, thereby becoming dependent on their partners' income and running a higher risk of descending into poverty.

Economic theory provides different accounts to explain the gender-specific patterns in labour division between couples. Family economic approaches identify structural differences in expected returns to labour market activity for men and women (in terms of wages, likelihood of promotion, etc.) as a key determinant for the households' decision on the concrete form of its labour supply – i.e., *who* is going to supply *how much* labour. From a policy perspective, the insights offered by economic theory suggest the following: If spouses imposed equivalent opportunity costs on the household by withdrawing from the labour market in favour of household production, then either spouse will do so with equal probability[2].

[2]Assuming, of course, that a withdrawal of either of them is still beneficial for the overall household welfare – i.e., purchasing household services externally imposes higher costs than one partner's (partial) labour market absence.

We would expect to find roughly equivalent opportunity costs when partners have comparable characteristics in terms of their education and experience and if the demand side for labour does not discriminate systematically against one sex. Over the past decades, many industrialized countries have made substantial progress with respect to these preliminaries, as e.g. steady rises in female educational attainment on the supply side and affirmative action policies on the demand side show (OECD, 2012). These improvements should level the opportunity costs that males and females impose on their common households when they abstain from the labour market. Yet, within the vast majority of families, we still observe a form of labour division where the women cut back from labour market participation (Bühlmann et al., 2010; Saraceno, 2011).

These observations inevitably lead one to question the accuracy of an economic analysis focusing on gender differences in expected labour market outcomes as the main reason for unequal labour division within couples. The answer to this question has important implications for equalizing-policy: If eliminating gender differences in expected returns to labour market activity (e.g. by raising female educational attainment, affirmative action, etc.) is not sufficient to ensure that couples' decisions on labour division will disadvantage one or the other partner with equal probability, the structural problem of female "underachievement" in the labour market will persist. Therefore, current policy approaches to improve female labour market outcomes may promise only limited success if men's and women's decisions on paid-labour participation differ depending on the social context – in this case, an individual versus partnership context.

The contribution of this chapter is to provide a direct, experimental test of this hypothesis. To the best of my knowledge, none of the few economic experiments concerned with couples' decision-making conducted so far focus on the participants' decisions on labour division.[3] It is a unique feature of the experiment

[3]Schröder et al. (2013) investigate the effects of individual vs. joint taxation on couples' labour supply, modelled as individual work effort. Cochard et al. (2013) explore how couples distribute resources when the initial allocation is determined exogenously vs. endogenously (i.e.,

presented here that it allows for a careful examination of how women and men divide paid and unpaid labour if objective measures on their individual productivity (i.e., expected pay-off), and hence potential gender differences in expected outcomes, are not available a-priori. Moreover, it allows for examination of how this decision changes within two particular social contexts: together with a stranger of the opposite sex or with one's real-life partner.

To achieve this, I recruited 20 real couples and 20 pairs of strangers and asked them to play a two-stage game; paired up either with their partner or a stranger of the opposite sex. In the first stage, participants make a joint decision on how to play the game: They can both complete a performance-based paid task (task A) or have one of the players perform an unpaid task (task B), thereby tripling the pay-rate for their partner playing task A. After completing their tasks, participants are informed about their pay-offs in private and then asked to make an individual decision about what proportion of their income to pay into a common pool, where it is increased by 20% and distributed equally between the two players. If couples maximize a joint utility function, or bargain cooperatively, they should be more likely than strangers to tolerate income-inequality and realise the efficient outcome (i.e., divide labour and play the combination A/B as opposed to each playing the paid task A individually). Furthermore, since neither men nor women know their productivity in either task, they should be equally likely to perform the unpaid or the paid task when dividing labour, irrespective of whether they cooperate with a stranger or with their partner. In order to verify that behavioural differences between familiar and unfamiliar participants cannot be attributed to a selection mechanism, I additionally collect a large number of personality trait and attitude measures that are typically thought of as driving factors for (gender-specific) differences in labour market orientation.

The remainder of this chapter is organized as follows: Section 3.2 sketches the theoretical accounts economic theory offers to explain the phenomenon of gender-

resulting from their individual work-effort). Beblo and Beninger (2010) document the attempt to investigate partners' provision of unpaid work vs. enjoying leisure-time experimentally.

specific labour division within couples. Section 3.3 contains a brief overview of the empirical, mostly experimental literature that revolves around family economics and labour division. Section 3.4 describes the methodology used for the experiments presented here, followed by Section 3.5, which provides the main results. Section 3.6 contains a summary of sensitivity analyses to verify the robustness of the findings. The discussion in Section 3.7 offers some potential interpretations of the results. Section 3.8 concludes.

3.2 Theoretical background

Economists have developed various theoretical approaches to model the family decision-making process, which can broadly be divided into two main classes, namely cooperative and non-cooperative models. Here, I will only briefly describe the two strands and how they relate to the experiment on labour division, which allows me to test some derivative model predictions. For an overview, consider for example Donni and Chiappori (2011); Grossbard (2011) and see Beblo (2001) for an application to bargaining over time allocation between partners.

Within the class of *cooperative models*, the allocation of time (to market and non-market production) was first modelled by Becker (1965) within a unitary household utility framework[4] and advanced by Gronau (1973a, 1977) and extended by Becker (1973) himself to a non-unitary framework. According to these models, the gender-specific intra-family labour division is optimal if, all else equal, wives

[4]This classification is not unambiguous, though. Grossbard (2011) argues that Becker's (1965) model does not rely on the assumption that the household maximizes a unitary (benevolent altruist's) utility function and may instead be regarded as an independent individual's utility maximization within a household, where partner income enters the constraints. While Becker (1965) himself is not explicit on the interpretation, Gronau (1973a), in his article advancing the Beckerian model, states his interpretation in the introduction: "This new theory has revived interest in the family as the basic consumption unit.".

expect lower returns to labour market activities relative to their husbands.[5] The theory thus suggests that spouses, since they are assumed to maximize their household's production (of utility), base their decision about who will cut back labour market engagement in favour of household work on the partners' individual labour market opportunities: The spouse who can expect a lower pay-off from labour market activity imposes lower opportunity costs on the household when giving up labour market work in favour of household work.

Following this rationale, the model predicts specialisation to occur whenever it yields efficiency gains. This for example will be the case, all else equal, when unequal conditions in the labour market promote productivity differences in paid labour between women and men, such that it is in both partners' best interests to allocate males' labour to the market and females' to the household. By using their individual comparative advantages, partners maximize their joint output. It follows that, if comparative advantages are not systematically related to gender, partners should be equally likely to specialize in one or the other production.

It is important to note that this perspective on the household as a production unit does not necessarily require a unitary utility function. According to Becker (1973), in equilibrium, for a certain man and woman to be married to each other only requires that their individual share of the jointly produced output cannot be improved if they were married to another person or remained single. It thus follows that the division of output may be unequal if the second best options of partners outside their union are unequal. Intra-household bargaining and collective approaches introduced by Manser and Brown (1980); McElroy and Horney (1981) and Chiappori (1988) model this subsequent intra-household allocation of resources in more detail and emphasize the potentially diverging interests between spouses but do not depart from the assumption that families realise efficient outcomes. In bargaining models household-related work is often assumed to display

[5]Strictly speaking, a productivity (dis-)advantage in labour market activities is not a necessary requirement. Becker himself claims a biologically determined comparative advantage for women in household-related work, particularly in child-rearing (see Becker, 1991).

an imposition both partners seek to avoid. The partner with the higher bargaining power will be able to confer most of this inconvenience upon her spouse and individual bargaining power within a relationship is again determined by a partner's outside options. Hence, gender-related differences in expected labour market outcomes will strongly influence the intra-household-bargaining process, especially with respect to labour division and allocation of resources.

The class of *non-cooperative models* of household decisions (first introduced by Lundberg and Pollak, 1994; Konrad and Lommerud, 1995), on the other hand, does not rely on the assumption that partners will realise the efficient outcome, since the threat point in case of disagreement is usually modelled not as the resolution of the union but as partners behaving non-cooperatively within it. Diverging interests within the household may thus lead to inefficiencies.

Although cooperative and non-cooperative approaches may differ with respect to the question *if* partners' agree to specialize, they yield similar predictions regarding *which* specific labour division arrangements they are most likely to choose: Lower expected returns from labour market activity for females either lead to a comparative advantage in household production or to a bargaining disadvantage in negotiations concerning who will be responsible for household work. Either way, when facing labour market inequalities to their disadvantage, women are thus more likely to reduce paid market activity in favour of unpaid household-related work. However, this need not be the case when household decisions are determined non-cooperatively (because couples simply may not divide labour if they do not agree) and, moreover, should not occur when expected returns from labour market activities are equal. The following section seeks to provide a brief overview of the experimental contributions made so far in an attempt to test which theory predicts couple behaviour regarding labour division the most accurately.

3.3 Related literature

So far we have established that, according to family economic theories, the decision on labour division between household members follows some kind of expected (joint or individual) utility maximization. Generally, a wide range of experimental evidence suggests that expected utility maximization might not be an accurate predictor of people's choices.[6] More specifically, violations of expected utility theory become particularly likely once an individual's decision affects others, and his outcomes are in turn affected by other actors.[7] This indeed applies to many, if not most, real-world decisions, especially within the household context.

Notably, an overwhelming majority of this evidence stems from economic laboratory experiments recording outcomes of strategic games played among strangers. Variations on the degree of information provided about fellow players show, however, how increasing familiarity with a partner affects outcomes (e.g. Eckel and Grossman, 1998; Ben-Ner et al., 2004; Cochard et al., 2009; Peters et al., 2004). Not surprisingly, the tendency to exhibit opportunistic behaviour seems to decrease while the willingness to cooperate increases with the tightness of social ties.[8]

Couples' decisions have been subject to a variety of experimental studies, since they are often subject to a trade-off between efficiency and equality. The majority of these studies focus on the unitary family utility model and aim to reveal its predictions to be inaccurate. The model's major shortcoming derives from its fail-

[6]For example, very prominently demonstrated by Kahneman and Tversky (1979).

[7]Illustrative evidence can be found in dictator games, ultimatum games, public good games, (see for example Thaler, 1988; Camerer and Thaler, 1995; Fehr and Schmidt, 1999; Fehr and Gächter, 2000a; Falk et al., 2008).

[8]For example, Peters et al. (2004) conduct experiments to investigate the behaviour of families in public good games. They ask participants to decide how much of their private endowment or pay-off they invest into a common pool; the amount collected is then multiplied by some factor greater than one and re-distributed in equal shares among all players, regardless of their initial contribution. The authors find that family members contribute higher shares (and hence generate higher overall pay-offs) when playing among themselves only, as opposed to playing in mixed groups with strangers. Cochard et al. (2009) demonstrate that, in symmetrical prisoner's dilemma games, 73% of participants cooperate when playing with their partners, as opposed to only 43% of those playing with a stranger of the opposite sex.

ure to acknowledge that interests and preferences within the household, in reality, may well diverge. Experimentally, this has been shown by, for example, Munro et al. (2008); Ashraf (2009); Iversen et al. (2011); Carlsson et al. (2013). Their findings suggest that couples may not simply pool their incomes, nor do they seem to make unitary decisions (let alone have homogeneous preferences) and maximize aggregate pay-offs—hence, they fail to reach the efficiency outcome predicted by the theory. Other authors' experimental findings provide evidence supporting this notion: Testing spouses' preferences for equality versus efficiency when choosing a pay-off distribution, Cochard et al. (2009) and Beblo and Beninger (2012) find their participants to prefer equality more often. In a recent study, Beblo et al. (2015) compared experimental results for German and French couples who were confronted with an equity-efficiency problem: Both groups displayed a significant inequality aversion, which was more pronounced among German couples.

Oosterbeek et al. (2003) study the conditions under which specialisation arises, an idea which is related more closely to the research question underlying this chapter. However, they examine anonymous interactions between randomly matched partners in a standard student subject pool. The authors design a bargaining game to mimic real world situations: Specialisation increases overall income, while reducing bargaining power for the partner specializing in household production (usually the wife). However, in the experimental set-up designed by the authors, participants choose these presumed *consequences* of labour division, not the labour division itself (i.e., a smaller pot combined with symmetrically distributed bargaining power vs. a larger pot inducing asymmetric bargaining power). Furthermore, since they play with completely anonymous partners, "real" gender effects are not subject to the authors' analysis. Instead, they focus on analysing a typical "wife decision problem" on a more abstract level.

Partially contradicting the standard game-theory predictions, Oosterbeek et al. (2003) find that "wife players" do actually choose the efficient outcome – even though this requires them to sacrifice bargaining power towards their "husband

player" – at least as long as the asymmetry it imposes on the individual bargaining power is not too large. The authors conclude that if the decision to specialize in household production at a personal cost is additionally framed in an affective relationship, the actual share of people willing to sacrifice their bargaining power in order to maximize aggregate welfare might be even higher in reality. However, they do not address the implications of the gender bias in this particular 'willingness to sacrifice' observable in real life where the majority of "these people" are actual women, not just wife *players*. The important insight their study offers concerns the question of whether people bargain cooperatively or non-cooperatively. The results suggest this to depend on the potential gains of cooperation relative to the degree of asymmetry in costs it imposes on the partners. Since they find that even completely anonymous strangers cooperate quite frequently when this asymmetry is relatively small, we might expect couples to accept even larger asymmetries in costs before they switch from cooperative to non-cooperative bargaining.

The experiments sketched above mostly present couples (or pseudo-couples) with decision and bargaining problems that involve the distribution of monetary pay-offs as such and do not explicitly focus on the decision of *how* to divide labour. This has, to the best of my knowledge, so far not been studied experimentally. Studies analysing gender-specific time disposal (a direct outcome of labour division within a couple) correspond to the underlying research question of this chapter more closely, but are mostly based on survey data. International European time-use data shows a gender gap in the average weekly workload for non-single parents in employment across virtually all EU-27 countries: Women work more (paid and unpaid work combined) and enjoy less leisure time (see Torres et al., 2007, 40). Contrary to the experimental studies on intra-couple income distribution sketched above, econometric studies exploiting time-use data rather support the notion of partners realising efficiency gains at the cost of equality. Generally, when couples are found to practice more asymmetric labour division, this is often interpreted as evidence in favour of the cooperative bargaining model.

Indeed, the gap in time spent on household-related activities appears to be influenced by the share women contribute to overall household income: For example, Beblo and Robledo (2008) show a woman's relative bargaining power to increase the more she specializes in labour market production. Bittman et al. (2003) and Haberkern (2007), using time-use data from Australia and Germany, respectively, show that women are able to reduce their workload in the household when increasing the share they contribute to the monetary income of the household, until these contributions are equal. But strikingly, when women contribute even more, their household-related workload increases again. The authors interpret this as the point where "gender trump(s) money" (Bittman et al., 2003), or, more precisely, a 'penalty' for violating the prevailing social norms. Indeed, this observation might point to some sort of cognitive bias, as in many developed countries modern couples may hold the ideal of gender equality as an abstract desirable goal in their minds but have not yet been able to incorporate it into their daily routines and habits (Fortin, 2005; Bühlmann et al., 2010; Miller and Sassler, 2010). The experiment described in the following section aims to determine whether this apparent bias can be observed in the lab.

3.4 Methodology

3.4.1 Hypotheses

The experiment described here aims to contribute to an important question arising from economic theories of the family. Do couples always realise efficiency gains when this requires specialisation in tasks? And are their specialisation patterns gender-neutral when comparative advantages are not related systematically to one sex? I intend to test the following hypotheses:

(1) Couples are more likely than strangers to agree on realising efficient outcomes (welfare gains) when this creates inequality (by requiring one player to give up income autonomy).

(2) When playing with their real partner, women are more likely than men to give up income autonomy in order to reach efficiency gains.

The first hypothesis is uncontroversial from a cooperative model perspective and previous experimental studies demonstrate how familiarity increases participants' willingness to cooperate (for example, as cited earlier Peters et al., 2004; Cochard et al., 2009). Non-cooperative models, however, may predict a different outcome, especially when partners perceive the asymmetry of individual costs from labour division as high (Oosterbeek et al., 2003), thus resulting in a higher probability of failure to reach the efficient outcome.

The second hypothesis is, however, clearly at odds with the predictions derivable from conventional family economic theory. According to the standard models, if cooperation occurs more often among familiar couples, then male and female partners should give up income autonomy (specialize in unpaid household-related work) with equal probability, assuming there is no comparative advantage for paid and unpaid tasks that is systematically related to one sex. But if familiar women are willing to perform an unpaid task more frequently than their unfamiliar counterparts, this will lead to an unequal distribution of independently controlled income within familiar couples. Female partners may be right to expect their partners not to exploit their advantage but to behave reciprocally instead, thus rewarding her for sacrificing her equal position deliberately (as documented in Oosterbeek et al., 2003). However, in terms of unconditional access to individual income, they would subsequently depend more on their partner's good will than unfamiliar females.

In order to test the first hypothesis, the act of performing a paid and an unpaid task must provide the unity of two players with a larger income than the pairs that

perform two paid tasks individually, thus representing the efficient outcome. In addition, the pay-out rules must reveal a-priori that realising the efficient outcome will generate income inequality among the two players.

To make the second hypothesis testable, the exact nature of the tasks must be unknown to participants. Consequently, partners should not anticipate gender differences in their expected pay-offs. More precisely, for the hypothesis to be rejected, women should not be more likely to undertake the unpaid task, regardless of whether their male partner is a stranger or their real partner.

3.4.2 Experimental design

In two different treatment groups, participants are paired up either with their partner or with a stranger of the opposite sex. They are asked to make two different decisions, at two different stages of the game. At the first stage, players must decide *jointly* if and how they want to divide labour. They have two real-effort tasks to choose from: Task A, a quiz which offers a performance-based pay rate for each correct answer and task B, an "assisting" task, that can be completed complementary to the paid task, but does not in itself yield any pay-off[9]. Instead, it triples the pay rate for the task-A performer. They can either:

(i) Work individually (both each spend ten minutes on task A, for individual performance-based pay-offs); or

(ii) Work together with their partners (one performs task A for a pay-rate while the other one completes task B to triple their partner's pay-off; however, only the task-A performer will receive a payment).

Throughout the decision process in step 1 of the first stage (see Table 3.1 for a chronological list of each step in the experiment) participants actually face each

[9]Participants had to type their partner's answers for task A from a paper-pencil answer sheet into a spreadsheet on a computer. The exact nature of either task is unknown to the participants. They are made aware, that both tasks involve real effort and that their completion is mandatory in order to generate a pay-off.

other and decide together whether, and how, to divide the tasks. Hence, partners in the control group do not know each other but are not anonymous. Only after they have reached a decision, partners are separated into different rooms, where they complete steps 2-4 in private. Thus, participants perform their tasks individually and afterwards decide privately how much of their personal income, if any, they want to invest in a common pool. This decision is of course conditional on the player performing the paid task A in the first stage and earning money.

Table 3.1: Course of the experiment

Stage 1	
Step 1	Decision 1 (jointly): Who does which task?
Step 2	Participants perform their "work"
Step 3	Participants receive their pay-offs in private
Stage 2	
Step 4	Decision 2 (individually): How much of their received pay-off do they want to invest in a common pool? (investments are multiplied by 1.2 and the resulting amount is split 50:50 for both participants)

The game and all of its stages were explained in detail before participants made any decision and a set of test-questions ensured that they understand the consequences of all choices available to them at any given point[10]. It is important to note that the exact nature of either task is unknown to the participants prior to their decision. They are solely informed that task A is some sort of quiz containing many different types of questions from a wide variety of fields, with the goal of solving as many questions as possible within ten minutes. Each correctly answered question yields a pay-off (which is tripled if one partner does task B). Task B, as participants were informed, is some kind of "assisting task", that does not require a certain level of performance and is solvable for anyone, but also requires effort and

[10]See Appendix 3.9, for the complete experimental instructions participants received.

must be completed in order to triple the partner's pay-off. This way, a priori gender biases should not evolve because participants cannot regress on any objective measures to estimate individual productivity. Hence, they should not be able to predict absolute and/or comparative advantages and divide the tasks accordingly. Therefore, *if* they divide the tasks, males and females should be equally likely to perform either task A or B in both the familiar and the unfamiliar condition.

A gender bias in the selection of tasks could still emerge, however, if the tasks were not "gender-neutral", i.e., if stereo-typical beliefs about one gender possessing a greater ability in performing a task exist (irrespective of the true ability distribution). It is therefore important to reflect on the implications for this study, if participants exhibit a pronounced bias in their a priori *beliefs*, e.g., if there was a stereotype that women, on average, are better quiz-takers[11]. If this was the case, we might observe women to be significantly more likely to play task A than half of the time, but this would hold constant regardless whether they play with their partner or a stranger, and the same would of course be true for men.

Furthermore, limiting prior information about the tasks prevents participants from estimating how many correct answers one could realistically score within the given time-interval. This ensures that the pay-offs remain private information to the individual generating it[12].

[11]Since the quiz was introduced as containing a wide variety of different questions, it can be claimed to be a rather gender-neutral task, as even subjects concerned with stereo-typical beliefs may have expected questions that are "typically easy for men but not for women" and those of the opposite type to be just as likely to occur.

[12]Whether pay-offs are public or private has been shown to have different effects in varying experimental settings with couples: In a field experiment conducted by Ashraf (2009) in the Philippines, men were more likely to store pay-offs in their personal accounts when they solely were informed about them in private. However, once an individual's pay-off was public information to both spouses, men were more likely than women to commit to pooled consumption. In a lab experiment conducted by Cochard et al. (2009), participants were asked to allocate tokens among themselves, with each partner having an individual exchange-rate that was private information. The authors found a clear majority of partners revealing their private exchange rate in the bargaining task and hence trying to realise efficient outcomes instead of using the chance to behave opportunistically.

Following standard economic game theory the following predictions derive: Via backwards induction, it becomes evident that rational players, when facing their last decision at step 4, have no incentive to invest anything into the common pool. This is a dominant strategy because it maximizes individual income for any given strategy of the other player. This holds, regardless of how their income was actually determined, i.e., whether the other player played task A or B. Therefore, at the preceding stage, a rational player would always choose to play task A, since she can anticipate the consequences of playing task B: This strategy will not yield any pay-off since a rational counter-player will not invest into the common pool. In short, standard game theory predicts that participants will never cooperate, neither at stage 1 of the game when they have to choose how to perform the task, nor at stage 2 when they have to choose an investment into a common pool. Hence, we should observe all participants playing task A and nobody investing in the common pool. However, we might observe couples cooperating if they pool incomes to maximize a unitary utility function or bargain cooperatively. Thus, observing spouses' behaviour at the first stage and comparing it to unfamiliar participants' decisions allows testing the first hypothesis.

The specific design of the game requires one player to be willing to deviate from this dominant strategy in order to maximize aggregate pay-offs. This involves a high risk, as it requires the player to give up control over his individual income, hence sacrificing his financial autonomy. In fact, players' willingness to cooperate is tested twice: At stage 1 when participants decide whether or not to cooperate by dividing the tasks, i.e., play either the combination A/A or A/B, and again at stage 2, when they must decide how much to invest into the common pool. Thus, it is possible for players to choose a form of cooperation that does not maximize aggregate welfare, but still increases it without requiring an a priori disadvantage of one player, i.e., both play the paid task A and invest their income (partly) into

the common pool[13]. If we observe couples to frequently choose this strategy, this would provide evidence in support of non-cooperative bargaining models.

3.4.3 Additional measures

Socio-demographic characteristics

After completing the game, participants fill out a questionnaire to provide basic socio-demographic information, including age, gender, family origin, socio-economic background, subject of study, duration of and satisfaction with their relationship (on a 10-point-scale) and relationship-related living arrangements and division of housework. In addition, the questionnaire contains an item to verify that participants in the unfamiliar condition did not know each other and that participants in the familiar condition were actual couples[14].

Measures for personality traits and individual attitudes

In addition to standard questions about socio-demographic characteristics, the questionnaire contains specific statements that gauge participants' degree of consent, thereby providing measures for certain personality traits, locus of control (LOC) and core self-evaluation as they are commonly applied. Furthermore, the questionnaire featured items that are typically used to elicit participants' taste for "challenge and affiliation". Further items address participants' attitudes on gender roles. All of these measures may be viewed as proxies for labour market preferences—in fact, a whole body of literature suggests that the gap in female and male labour market performances can be linked to differences in preferences (for an overview and critical examination see Trzcinski and Holst, 2011). Accordingly,

[13]It is obvious, however, that a disadvantage may still arise, if players do not invest equal shares or if one partner performed worse in the quiz and therefore simply has less money at his disposal to invest.

[14]Participants were asked to state their partners' birthday, which you of course are much less likely to know by heart if you are not involved in a romantic relationship with that person.

evaluating whether these variables are related to certain specialisation patterns is crucial to this study.

3.4.4 Treatment groups and participants

Eighty people participated in the experiment. Participants were mainly recruited among the University of Warwick student body. The game was played in two different treatment groups, with individually scheduled sessions for each of the 40 pairs:

- Heterosexual couples
- Pairs of strangers, mixed-sex

Participants were predominantly graduate students (53% Masters; 13% PhD) and under-graduate students (28%); 8% of participants[15] claimed not to be enrolled as a student at the time of the experiment. Participants were recruited via advertising (posters and flyers) on campus.[16] The distribution of participants over study levels varied only slightly between the two treatment groups, with the unfamiliar participants comprising a larger share of Master students and the familiar group representing a relatively larger share of PhD students. The share of undergraduate and non-students is equivalent in both groups. The average age of participants was 25.[17]

Participants in the familiar group by definition are all involved in a relationship. However, participants in the control group, although unfamiliar with their experimental partner, are not necessarily single. In fact, 30% of female and 25% of male participants in the unfamiliar group reported being in a relationship. These

[15]May not add up to 100 because of rounding.

[16]Couples in the treatment group and unfamiliar individuals for the control group were recruited via separate advertisements.

[17]The exact statistics: M=25.10, SD=4.49. The fact that the sample consists of 92 % university students who were largely in their mid-twenties should necessarily be born in mind when deriving conclusions. See section 3.6 for a more thorough discussion.

compositional differences between the groups are addressed in section 3.6, which provides a detailed analysis of potential selection threats to the robustness of the results.

3.5 Results

In the following section, I use the collected experimental data to evaluate the stated hypotheses by answering the following questions: Are familiar couples more likely to cooperate at the cost of equality and thereby able to realise greater joint outcomes? Are women more likely than men to give up their individual, independent income when they play with their real partner? Are the gains in aggregate welfare for familiar couples therefore primarily realised at the expense of female income autonomy?

3.5.1 Hypothesis I: Couples are more likely than strangers to realise efficient outcomes.

Table 3.2: Proportion of Participants Cooperating by Stage and Familiarity

	Familiar	Unfamiliar
Stage 1: cooperation (specialisation: A/B)	100% (n=40)	60% (n=40)
Stage 2: partial cooperation (non-specializing A/A players pooling income)	— (n=0)	62.5 % (n=16)

Read: In the familiar group, all couples (100%) cooperate by dividing labour (i.e., play the combination of tasks A/B) in the first stage. In the unfamiliar group 60% of participants divide labour, i.e., 24 out of 40 participants. In the second stage, out of those people who have not divided labour but instead performed the paid task individually, 62.5% cooperate by investing their stage-1-income (partly) into a common pool.

Table 3.2 shows the proportion of people cooperating at the different stages. At stage 1, the number of people who specialize by dividing the tasks and play the game as A/B performers were 40 in the familiar and 24 in the unfamiliar group.

Thus, all familiar participants cooperate, but "only" 60% of unfamiliar players.[18] This difference is statistically significant.[19]

As discussed above, welfare gains can only be reached by choosing a division of labour that requires one player to give up control over his personal income and allows the other player to determine their final pay-off (recall that the task-A player alone receives a pay-off at the end of stage 1 and thus is the only one to decide about how much to invest in the common pool at stage 2, i.e., task-A players determine both their own and their partner's final pay-off). Presumably, participants will only be willing to perform the unpaid task B when they expect their partner to behave reciprocally by investing their pay-off in the common pool, thereby sharing the fruits of their labour.[20]

Another form of "partial" cooperation evolved among unfamiliar players and is noteworthy. As shown in the second row in Table 3.2, of the 16 players who did not cooperate at the first stage, i.e., where both partners completed task A, 10 invested their entire income into the common pool[21], which can be interpreted as an attempt by the players to cooperate while sustaining individual control over their personal incomes, yet, within this constraint, trying to maximize aggregate

[18]Compared to the standard game-theoretic predictions, this might actually be viewed as a surprisingly high rate of cooperation among strangers. This can be viewed as a form of a trust game, where even completely anonymous players have been recorded consistently to cooperate by "trusting" (Berg et al., 1995; Fehr and Gächter, 2000b; Oosterbeek et al., 2003). The fact that most participants shared a common identity as students could have driven up the cooperation rate. Furthermore, even though participants were assured that their income and their investment decision would be kept secret from their partner, it was obvious that at least to the experimenter, they were known instantly – which might have also favoured the high investment rate and the small rate of opportunism in the unfamiliar condition.

[19]$\chi^2(1)=10$, $p=.001$.

[20]Among co-operators in both groups, however, two task-A-players (roughly 10% of familiar and 17% of unfamiliar co-operators) did not fulfil their part of the deal to the full extent and exceeded opportunism: i.e., those "defectors" invested only a share of their stage 1 earnings. Although this type of opportunistic behaviour approaches the homo-economicus behavioural predictions, none of them let their partners down completely. The minimum invested was 49% of the amount earned in task A among familiar couples and 60% among unfamiliar cooperators.

[21]80% of them actually managed to coordinate, i.e., both partners mutually invested all their income.

welfare.[22] This can be interpreted as a form of cooperation that favours equality of partners over the efficiency of their joint outcome.

Based on these figures, the first hypothesis cannot be rejected. Familiar couples seem to strictly prefer efficiency over equality.

3.5.2 Hypothesis II: Women are more likely than men to give up income autonomy in real couples.

Table 3.3: Number of People Performing Task A and B by Familiarity and Gender

	Familiar		Unfamiliar	
	Male	Female	Male	Female
Paid-task-performers (A)	14	6	13	15
Unpaid-task-performers (B)	6	14	7	5
n =	20	20	20	20

Read: In the unfamiliar group, 13 out of 20 males perform task A.

The first row in Table 3.3 shows the number of males and females performing the paid task A (of all participants in their treatment group). In the unfamiliar condition, 15 out of 20 females completed task A, i.e., 75%. When playing with their partners, females are much less likely to do so, as only 30% of all familiar women perform the paid task. This difference is highly significant[23] and partly due to the fact that couples choose to specialize more often, i.e., the familiar condition overall has fewer task-A performers. Males, however, are not more likely to complete task B when playing with their female partner as opposed to a female

[22] Another possible explanation, which is rather speculative at this stage of research, involves male ostentation: in particular, males might feel the desire to impress their female partner by signaling they performed well in the task rather than potentially being suspected to not have generated much money to invest into the pool in the first place due to poor performance on the quiz.

[23] Fisher-exact-test: $\chi^2(1)=8.12$, $p=.004$.

stranger. Hence, they act as task-A performers in both groups about two thirds of the time, tests indicate no significant difference between the conditions. This implies that couples' higher likelihood to divide labour derives from women's greater willingness to perform the unpaid task when playing with their partner. We can verify this by looking only at those participants who choose specialisation.

The second row depicts the behavioural pattern of participants cooperating at stage 1, i.e., they play the combination of task A and B. For familiar participants the distribution is symmetric, as all of them cooperated at the first stage. Thus, familiar male and female task-A performers (and task-B performers, respectively) total 20. Among unfamiliar participants, there are generally more task-A performers than task-B performers, because not all of them cooperate with their partners. The number of unfamiliar male task-B performers reveals what proportion of the 15 unfamiliar female task-A performers where co-operators: Since 7 men performed task B, by definition, 7 women out of the number who performed task A were their cooperating partners (and vice versa).

Familiar females perform the unpaid assisting task B in 70% of all cases, whereas when cooperating with strangers in the unfamiliar condition, less than half (only 42%) of females perform task B. Economic theory suggests, however, that they will perform either task with equal probability in the absence of a comparative advantage. That is, once they decide to cooperate with their partners, females and males should be equally likely to perform the unpaid task. This should hold regardless of whether they cooperate with a stranger or their partner. As a test of given proportions reveals, the theoretical predictions match the actual decisions of unfamiliar cooperators very accurately: the probability does not differ significantly from one half. When cooperating with their partners, however, familiar females' probability to perform the unpaid task B is significantly higher than .5.[24]

[24]The exact test-statistic for familiar females is $\chi^2(1)=3.2$, $p=.037$ against the one-sided alternative that the probability of performing the unpaid task is greater than 0.5. For unfamiliar females, testing against the same one-sided alternative delivers $\chi^2(1)=.077$, $p=.609$.

3.5.3 Implication: Higher (gendered) inequality among familiar couples

If couples' higher co-operation rate is driven by females greater willingness to perform the unpaid task B, then by definition, they sacrifice their income autonomy more often. In order to quantify the implications of this finding, one may look at the generated pay-offs conditional on participants' specialisation and pooling decisions. Recall that by cooperating at stage 1 (playing the A/B combination), participants can triple their pay rate per correct answer. However, only one of the partners is performing the task and hence collecting the pay-off. By cooperating at stage 2, the accumulated earnings can be increased by yet another 20% (the mark-up factor α and will then be split equally between both players. The overall pay-off at the end of stage 2, π_{2i}, for a player i, therefore depends on her own contribution c_i (the share of stage-1-income, s_i, invested), and that of her partner j, given their individual stage-1 pay-offs ($\pi_{1i,j}$):

$$\pi_{2i} = \pi_{1i} - c_i + \alpha(\frac{c_i + c_j}{2}), \text{with } c_i = s_i \times \pi_{1i} \tag{3.1}$$

This is a standard public-good game. The initial endowment π_{1i} over which a player decides is endogenous, since it depends on her performance x_i conditional on playing task A and on her pay rate r_i, which is determined by whether or not her partner j also performs task A or instead plays the assisting task B. An individual's stage-1 pay-off is therefore given by:

$$\pi_{1i} = x_i(A_i) \times r_i(A_j) \tag{3.2}$$

Hence, stage-1-income is zero for all players who play task B. Among those who play task A, assuming performance is constant, players whose partner is willing

to perform task B receive a three times higher pay-off. In stage 2, the pay-off depends on the share of income that players i and j invest into the common pool. Task-A players who play together with a B-task playing partner know that they are the only ones to invest in the common pool because their partners receive no income from stage 1.

Table 3.4: Simulated earnings after stage 1, by familiarity and gender

	Aggregate	Male	Female	Difference (M – F)
Familiar	4.05	5.67 (3.81)	2.43 (3.81)	3.24***
Unfamiliar	3.51	3.01 (3.19)	3.92 (3.33)	.81
Difference (F – U)	.54	2.57**	-1.48**	

Note: Given participants' actual specialisation decisions, earnings are simulated at a constant performance-rate of 9 correctly scored questions. This corresponds to the average of familiar males' actual performance. Mean earnings in £; standard deviation in parentheses. Differences in means: significance indicated at * 10%, ** 5%, *** 1% level.
Read: In the familiar group, males on average earn 5.67£ after the first stage. Females in this group on average earn 3.24£ less, a mean value of 2.43£ precisely. Compared to familiar women, unfamiliar women on average earn 1.49£ more after the first stage, generating a mean income of 3.92£.

Table 3.4 shows the simulated average incomes for the two different groups after stage 1, given the participants' actual decisions. Holding performance constant at 9 correctly scored questions[25] for every task-A player allows us to examine how players' specialisation and pooling decisions affect the distribution of income. Stage 1 earnings reflect the different decision-patterns regarding specialisation. The aggregate difference between the familiar and the unfamiliar group in stage-1 earnings is £0.54 and is not significant. A closer look at the distribution in stage-1 earnings by familiarity and sex in Table 3.4 reveals the gendered labour division dominating in the familiar group. While no significant gender gap in stage-1 earnings can be found in the unfamiliar group, familiar females on average earn

[25]For an evaluation of participants' actual performance by groups, please refer to the robustness checks provided in section 3.6.

£3 less than familiar males. Another consequence of these specific specialisation patterns manifests in the gaps within sex by familiarity. While familiar men realise incomes which, on average, are roughly £2.50 greater than incomes generated by unfamiliar males (because they are more likely to reap efficiency gains), familiar females, on average, earn about £1.50 less than their unfamiliar peers (because they are more likely to give up their income autonomy and not earn an income at all).

Table 3.5: Simulated earnings after stage 2, by familiarity and gender

	Aggregate	Male	Female	Difference (M – F)
Familiar	4.82	4.21 (.52)	3.89 (.52)	.32**
Unfamiliar	3.99	3.10 (1.09)	3.79 (.90)	-.69*
Difference (F – U)	.83***	1.11***	.10	

Note: Based on the simulated earnings for stage 1, stage-2 earnings are simulated given participants' actual investment decisions. Mean earnings in £, standard deviation in parentheses. Differences in means: significance indicated at * 10%, ** 5%, *** 1% level.
Read: In the familiar group, males on average receive £4.21 after the second stage. Unfamiliar males earn a mean value of £3.10

Table 3.5 shows that the gender differences vanish after task-A performers reward their task-B-performing partners at the second stage: By investing their income in the common pool, A-players increase it by 20% and share it equally with their partners. Since nearly all cooperators[26] invest their complete income, at the end of stage 2, reciprocity has smoothed out the variance in income established at stage 1 and differences in earnings between men and women within the familiar and unfamiliar group become negligible. As a result, familiarity remains the only factor to explain the variance in earnings, since it corresponds with a higher cooperation rate in the first place and since co-operators are more likely to invest their full

[26] As noted earlier, there were two exceptions both among familiar and unfamiliar co-operators, where a task-A performer was assisted by his partner (i.e., a task-B performer) and did not invest the entire sum earned.

earnings into the common pool (where they are again increased by 20%) than non-co-operators. From Table 3.5, it also becomes evident that the aggregate difference between familiar and unfamiliar participants' final earnings are driven by men. Male participants enjoy significantly higher terminal earnings when playing with their female partner as opposed to men who play with a stranger (£1.11, a mark-up of roughly 35%). Thus, they are able to reap the benefits from specialisation. For female participants, surprisingly, playing with their partner does not yield an advantage over playing with a stranger in terms of the final pay-off generated.

3.6 Robustness checks

The validity of the results presented relies crucially on the assumption that participants in both groups, apart from the differential treatment they receive (playing with their partner or playing with a stranger), do not differ with respect to other characteristics that might influence their decisions. This is basically identical to claiming that familiar females would behave just as unfamiliar females if they played with a stranger. Therefore, the main concern is whether those females playing with their partner differ systematically in some important characteristic(s) that in turn make them inclined to choose the assisting task more often. If this were the case, the results would likely suffer from selection bias. This section offers a closer examination of potentially confounding variables, in order to mitigate apprehensions in this regard.

3.6.1 Performance

Since the findings of this experiment record differential decisions on specialisation for familiar and unfamiliar participants *despite the lack of objective measures to predict comparative advantages*, the first concern relates to actual productivity: The average number of quiz questions solved should not differ for men and women within or between both groups.

Table 3.6: Performance by familiarity and gender

	Aggregate	Male	Female	Difference |M – F|
Unfamiliar (n=27)	6.11 (3.23)	6.23 (3.11)	6.00 (3.44)	.23
Familiar (n=16)	8.06 (5.01)	9.09 (4.93)	7.17 (4.02)	1.92
Difference |U – F|	1.95	2.86	1.17	

Note: Mean correct questions given by task-A performers; standard deviation in parentheses. All differences in means are tested with a Mann-Whitney test – none of the differences show statistical significance below the 10% level.
Read: In the unfamiliar group, participants on average scored 6.11 correct questions, with familiar males scoring a mean of 6.23 and familiar females 6.00.

Table 3.6 summarises the average number of correct answers participants gave when performing task A, which overall range from 0 to 16.[27] The most important observation is that differences in participants' performance do not differ significantly for any group or sub-group comparison. Despite the lack of significance, by examining the table at face value, one may still be worried by familiar males' relatively high performance. This may well be related to the fact that the vast majority of familiar men play the role of "provider" and thus might simply increase their effort since they have to earn income for two people. And indeed, when test-

[27]The results of four participants had to be excluded for calculating the means. They admitted (and their answer sheets also proved this) to have "cheated", all of them in the same way: They knew it was impossible to solve all questions within the given time interval of ten minutes (this was public information), so they reserved the last minute of their "work time" to randomly guess the multiple-choice answers to those questions they had not yet answered. This was not explicitly prohibited, so strictly speaking they were not cheating. However, by doing so they were able to solve presumably roughly as many questions as other participants plus the extra share scored correctly by chance (wrong answers did not affect income; this was public information, too). I am able to identify the participants in question (because, during the debriefing, they admitted to have applied this strategy) and I can also be sure that this was not the case for any other participant (as their answer sheet would have revealed such a strategy even if they had not told me). However, I cannot identify exactly how many questions "cheaters" were able to "honestly" solve and how many they simply guessed correctly. Therefore, I am unable to correct their score, which is why I decided to exclude them completely from the analysis of the participants' performance. Three of these cases (all male task-A players) occurred in the familiar group, and one (a male task-B player who "added" guessed answers to his partner's multiple-choice-answers when copying them into the spreadsheet) in the unfamiliar group.

ing the cooperators' performance (across both groups) against non-cooperators', I find a significant difference: Cooperators on average score 2.36 more answers correctly.[28] This is consistent with other experimental studies demonstrating a positive effect of higher piece-rates on performance (for an overview see Dohmen and Falk, 2011).[29].

3.6.2 Trust level

Perhaps not surprisingly, I find familiar and unfamiliar participants to differ substantially in their average reported trust level: Paired up with their partners, players report significantly higher trust (M=9.12; SE=2.27) than unfamiliar partners (M=5.89; SE=2.65)[30], on a scale of 0-10, where 0 represents not trusting one's partner at all. However, no effect of gender on the trust levels can be observed, and, more specifically, no interaction between gender and familiarity—i.e., the increase in trust when playing with one's partner as opposed to playing with a stranger does not differ for females and males—which rules out trust as a potential explanatory variable that could account for the difference in familiar and unfamiliar females' behaviour. However, it is possible that an increase in trust toward one's partner, even if it is quantitatively the same, influences women's behaviour in different ways than men's.

3.6.3 Differences in attitudes and personality traits

Among the various personality and attitude measures collected, very few significant differences were found, neither between sexes nor between unfamiliar and familiar

[28] W=1666.5, p=.05. Moreover, it is important to note that, among cooperating task-A players, performance does not differ significantly by gender.

[29] I have conducted further tests: Recalling the descriptive statistics provided on participants in 3.4.4 one could suspect that the higher share of PhD students in the familiar group might pose a problem in terms of productivity differences. However, testing the mean scores of PhD students against other participants' also confirms no significant differences in average performance.

[30] W=1183.97; p<.001

participants. Table 3.7 summarises the measures and focuses attention on the same-sex comparison of familiar and unfamiliar participants, in order to examine whether familiar females display a selection: One can easily see that the means do not differ significantly for familiar and unfamiliar females in any of the tested characteristics.[31]

This is important to highlight for two reasons: (1) The lack of significance in personality and attitude measures is very relevant in supporting the claim that females in the treatment group who played with their partners do not form a special selection. (2) Personality trait measures have recently gained in popularity for explaining (gender) differences in labour market outcomes (see for example Groves, 2005; Borghans et al., 2008; Cobb-Clark and Tan, 2011). The fact that they do not seem to govern participants' decisions on labour division in this experiment also emphasizes that they should be treated with a reasonable degree of caution. Some studies partially ascribe the gender gap in labour market performance to a self-selection driven by differences in personality traits, but they might very well display a *result* of gendered labour division instead (compare also the critical examination of reversed causality between labour market outcomes and locus of control by Trzcinski and Holst (2011)). At least in the study described here, participants did not exhibit any significant differences in the personality trait measures that are often assumed to determine preferences for labour market activity (such as locus of control, need for challenge or affiliation, traditional gender role attitudes). I will therefore briefly describe what these measures intend to capture.

Gender Role Attitudes

Participants were asked to indicate their agreement to the statement "It is a man's duty to earn money while the woman takes care of household and family" on a four-point scale (strongly disagree – strongly agree). If women in the treatment group

[31] Again, I have conducted further tests to confirm that there is no significant interaction effect between gender and familiarity that could explain the difference in the behaviour between unfamiliar and familiar women.

represented a selection of females who prefer traditional gender arrangements, we would expect them to agree more often with this statement. However, this is not the case. One might then hypothesize that this mechanism could work indirectly through their male partners who might, if they have more traditional attitudes towards gender-roles, subtly pressure their female partners into playing the assisting task B. However, the same comparison for males reveals, that they do not differ significantly in their average agreement with the statement either .

Locus of Control

Locus of control (LOC) is a psychological measure that intends to capture how much a person believes they are able to actively influence the course of and the events in her life. More precisely, the construct comprises two measures: The external LOC is an index of items[32] that gauges whether a person considers his life to be governed *externally*, i.e., a high external LOC ostensibly means that a person judges his own ability to exert influence in his life to be very limited. The internal LOC is an index constructed, correspondingly, from items intended to capture the opposite view, i.e., a person considers her life is governed *internally* (items (6)-(9) in Appendix 3.9). Thus, a high internal LOC supposedly coincides with the perception that life courses and events are mainly determined through one's own actions and decisions. Following these definitions, one might hypothesize that females who select themselves into a relationship are more likely to exert a higher external LOC, or a lower internal LOC, respectively,[33] and therefore are more likely to avoid responsibility (e.g., providing for themselves and their partners by performing task A) and instead try to delegate it to their partners. However, I again fail to detect significant differences between men and women, familiar and

[32]In particular, it equals the sum of scores assigned to five different statements (items (10)-(14) in Appendix 3.10).

[33]Precisely this constellation, a high external and a low internal LOC, is often hypothesized to be responsible for lower labour market outcomes of women, for an overview see Trzcinski and Holst (2011). I will get back to this point in the discussion provided in Section 3.7.

unfamiliar partners, or between the subgroups. This holds true not only for testing the indices (as presented in Table 3.7) but also when testing each individual item within the index.

Challenge and Affiliation

Two measures that are often linked to labour market success are "challenge" and "affiliation" indices. In general, people who score high on the challenge items are thought to have a higher drive for achievement and are hence more career-oriented (i.e., they find it important or very important to "accomplish something worthwhile" and to have "the chance at getting a promotion or a better job"[34]). People who score high on the affiliation items are assumed to be more agreeable and have a higher need for affiliation (they tend to rate "the friendliness of the people one works with" and "the respect of other people" important or very important[35]). Again, one could speculate whether familiar and unfamiliar females differ with respect to these characteristics, such that familiar females are less challenge-seeking than unfamiliar females relative to their partners and/or more affiliation-seeking and therefore prefer to "assist" their male partners more often rather than "perform" themselves. Again, surprisingly, no significant differences among the groups can be detected in the sample.

Big Five & Self-confidence

The "standard" personality measures that aim to quantify the degree to which a person exhibits certain character traits are the so-called "big five". A number of behavioural researchers ascribe substantial explanatory power to predicting a wide variety of outcomes to these measures, such as happiness, health, and especially labour market outcomes (for an overview see for example Judge and Hurst, 2007; Borghans et al., 2008). However, as in the case of locus of control, most studies

[34]Items (15) and (16) in the questionnaire, see Appendix 3.10.
[35]Items (17) and (18).

have not been able to address reversed causality issues adequately (Trzcinski and Holst, 2011).

Apprehensions of familiar females displaying a certain selection—e.g. because women in a relationship may display systematically lower levels of self-confidence and thus be more likely than unfamiliar females to estimate their own ability, i.e., productivity, as inferior to their partner—are not supported by the data. In particular, the items addressing participants' self-confidence, i.e., the statements "I am confident I get the success I deserve in life", "Sometimes when I fail I feel worthless", and "I am filled with doubts about my competence", warrant closer examination. Yet again, there are no significant differences between the female groups (and also not in comparison to their male partners, not shown). There are some small differences between familiar and unfamiliar males: familiar males are, on average, less likely to feel depressed and to feel worthless when failing; and they are more likely to be confident to get the success they deserve in life. This might be a potential mechanism that calls for further research. However, these results certainly do not support the hypothesis that familiar females display a particularly under-confident selection and hence shy away from the paid task.

3.6.4 Selectivity of the student sample

Further concerns might derive from the selectivity of student subjects who may be viewed as not representative of the "true" couple population. However, despite the standard reserves toward student samples used in economic experiments (for a thorough discussion, see Harrison and List, 2004), in the special case of the experiment presented here, the selection may arguably strengthen the results. The major concern towards student samples is usually that it disproportionately represents very young and highly educated individuals. In case of the research question underlying this study, however, this particular over-representation might actually strengthen the results: While I examine the behaviour of a selection with a pre-

sumably very high career- and labour market orientation, I *still* find gender-specific labour division.

Besides age and education level, the couples in the sample are also certainly not representative of the whole population of couples in terms of relationship duration. Almost half of all familiar couples were not (yet) cohabiting, and many had not even been together for a year.[36] It thus seems fair to assume that most of the participating couples had not yet established a sound partnership (in the sense that most of them did not live in a common household). This supports the notion that, if anything, the gender effect I find might be biased downward, since for these couples, behaviour might actually be driven by relationship-specific gender stereotypes in habits and routines to a lesser extent than in the "true" underlying population of all heterosexual European couples.

[36]Precisely, half of all familiar couples reported a time-span of 19 months or less when asked for the duration of their relationship.

Table 3.7: Attitude and personality trait measures

Trait or attitude measure	gender	familiar	unfamiliar	F–U
Traditional gender role attitude				
	Male	1.6 (.68)	1.4 (.6)	.2
	Female	1.3 (.47)	1.45 (.76)	-.15
Locus of control (ext.)				
	Male	12.75 (1.52)	12.15 (2.39)	.6
	Female	13.2 (2.21)	12.9 (2.53)	.3
Locus of control (int.)				
	Male	10.85 (1.87)	10.9 (2.14)	-.05
	Female	11.3 (1.66)	10.5 (3.09)	.8
Challenge				
	Male	6.68 (1.16)	6.2 (1.32)	.48
	Female	7.1 (.72)	6.8 (.95)	-.3
Affiliation				
	Male	6.95 (1.77)	7.05 (1.0)	-.25
	Female	7.45 (.89)	7.25 (.91)	.2
BIG 5				
Confidence in success				
	Male	3.35 (.67)	2.95 (.51)	-.4**
	Female	3.3 (.66)	3.32 (.47)	-.02
Feeling depressed sometimes				
	Male	1.8 (.70)	2.45 (.94)	-.65**
	Female	2.65 (.88)	2.6 (1.39)	.05
Feeling worthless when failing				
	Male	1.95 (.60)	2.45 (.89)	-.5**
	Female	2.2 (1.51)	2.32 (.89)	-.15
Doubts about own competence				
	Male	2.00 (.86)	2.35 (.88)	-.35
	Female	2.2 (1.01)	2.47 (.7)	-.27
Determining events in own life				
	Male	3.35 (.59)	3.25 (.79)	.1
	Female	3.1 (.97)	2.75 (.97)	.35

Note: Group means for 4-point scale answers (standard deviation in parentheses), where a higher number indicates a greater tendency to agree with or (in case of challenge and affiliation) to rate a given item as important. Locus of control and challenge and affiliation are indices containing several items, see 3.6.3 for details. All differences in means are tested with a Mann-Whitney test. Significance indicated at * 10%, ** 5% and *** 1%- level.
Read: Familiar males' mean answer to the statement "It is a man's duty to earn the money, while the woman takes care of household and family." is 1.6, which means that, on average, they stated to "disagree" with the statement slightly but insignificantly less often than unfamiliar males (1.4 mean).

3.7 Discussion

The observed difference in female behaviour when playing with a stranger as opposed to playing with their real partners does not seem to be driven by differences in individual personality and attitude characteristics. Thus, it seems unlikely that the results are driven by a self-selection problem in the samples. Even the level of trust, which increases substantially when comparing familiar and unfamiliar partners, does not show any variation by sex within the familiar group. Of course, it is possible that trust affects male and female behaviour differently, but further research is needed to verify this. Qualitative structured follow-up interviews could provide a fruitful way to explore participants' motives and the driving factors for their behaviour.

Although an analysis that merely relies on non-parametric tests of group means is necessarily limited, the fact that women in the familiar and unfamiliar group do not differ in means with respect to personality trait measures, is important to highlight. Familiar women are more likely to play the unpaid task than their male partners *even though* the design of the two tasks and their presentation did not provide them with objective measures to infer comparative advantages. At the very least, as the comparisons of personality and attitude measures to unfamiliar females show, they should not have more reason to assume productivity differences relative to their male partners.

If men and women in the familiar condition do not differ in their characteristics, neither with respect to personality traits, labour market orientation, nor gender role attitudes, how can the difference in their behaviour be explained? One possible explanation is gender priming. Several studies demonstrate how participants identify with gender-stereotypes when they are cued (even subtly) and, often subconsciously, "adjust" their behaviour (see for example Sinclair et al., 2006), even in the presence of economic incentives (Günther et al., 2010; Schmitt, 2013). If participants, when confronted with a stereotype, show a greater tendency to exhibit

behaviour consistent with that stereotype, the question becomes: Were familiar players more likely to be affected by "priming" than unfamiliar players?

Indeed, by construction of this experiment, a priming effect might have been at work: It was inevitable to reveal to familiar couples that their relevant characteristic qualifying them as study participants was their relationship with each other. Even though this background may not have made them consciously aware of a gender-related research question (many of them, as it turned out during the debriefing, believed it was concerned with their cooperation and opportunism strategies), it might still have imposed a much stronger cue to activate their identity as "man" or "woman". This may have caused a tendency to behave according to the stereotype in an effort (whether conscious or unconscious) to comply with social norms. Participants who were recruited for the control group, on the other hand, could not know in advance that the research question was in any way related to gender or couples. Many of them openly admitted that they had not even considered the possibility that I was looking into *how* they cooperate. Eventually, most of them did not even pay much attention to the fact that they were playing with a partner of the opposite sex.

From an economics perspective one might argue that non-conformist behaviour when facing a stereotype is costly for the individual and hence it might be perceived as optimal to behave in line with the stereotype, as long as the costs (in this case: financial independence within the experiment) do not exceed the costs for acting against the stereotype (here: a woman taking on the "provider role" within the experiment). Akerlof and Kranton (2000) have modelled such costs as "identity-utility"-loss. The studies of Bittman et al. (2003), Haberkern (2007) and Beblo and Robledo (2008) cited earlier provide empirical support for this notion: the desire to comply with gender-specific social norms in heterosexual households may indeed have measurable effects on how males and females divide household work. Women who violate the ruling social norms by contributing a larger share to the

household's income than their male partner must "pay" a penalty for their violation by also (re-) increasing their share in housework.

Other examples supporting the notion of costs for non-stereo-typical behaviour or deviating from ruling social norms can be found in the literature on divorce. For example, Amato and Booth (1995) provide further insights on why and how conformity with social norms might be beneficial for partners: With a longitudinal survey dataset from the U.S., they show that for women, changes from a traditional gender role attitude to a more progressive one coincide with a decline in their marital satisfaction level, whereas for men, the effect works in the opposite direction. Cooke's (2006) findings establish a link between the family model favoured by a country's policy at the macro-level, the practiced gender equality in formal and informal labour division at the micro-level, and divorce rates. Comparing a country where policy favours the male bread-winner model (Germany) to a country where national policy does not actively promote a specific family model (U.S.) allows her to draw the following conclusions: Couples deviating from the politically and institutionally supported family model in Germany (i.e., practice *more* gender equality) run a higher risk of divorce. On the contrary, in the U.S., relationships were more stable when labour division between spouses took on a more egalitarian form.

Interpreting these results relative to the findings in this experiment offers two (possibly complementary) explanations: Women in the familiar condition might (subconsciously) expect some form of 'penalty' for behaviour perceived as non-conforming with social norms and stereotypes. For example, in line with the argument put forward by Amato and Booth (1995), one such penalty for non-compliance might arise via a threat on females' individual level of satisfaction with their relationship. Following Cooke (2006), the perceived threat when not behaving according to the prevailing social norms might (alternatively, or even additionally) stem from an increased risk for the relationship's failure.

Examining perceived threats to the relationship or satisfaction within the relationship could ostensibly help assess men's and women's motives when dividing the tasks. These motives, however, can hardly be elicited by a standardized, anonymous questionnaire alone, which once more highlights the potential benefits of incorporating qualitative methods into the experiment. If social norms drive the differences in behaviour, then in the current design, it is almost impossible to capture these experimentally. Therefore, further investigation of this topic in the lab should integrate structured, qualitative follow-up interviews.

3.8 Conclusion

In this chapter, I presented experimental findings on specialisation decisions and labour division between partners (20 heterosexual couples and 20 pairs of strangers) who played a two-stage game. Paired up either with their real partner or a stranger of the opposite sex, participants were asked to make a joint decision on how to play the game in the first stage: They had to choose whether (i) both would complete a performance-based paid task (task A); or (ii) one of them would perform an unpaid assisting task (task B), thereby tripling the pay-rate for the task-A player. In the second stage, after completing their tasks, each participant was informed about her payment in private and asked to (iii) make an individual decision on investing her income (partly) in a common pool, where it was increased by 20% and then split equally between the two players.

The main results, in short, are: (1) All familiar couples cooperate, i.e., they play the game in the A/B combination. In the control group, a considerable share (60%) of unfamiliar participants cooperate in the same manner, while the rest chooses to play the combination A/A. For familiar couples, their greater willingness to cooperate at both stages rewards them with higher overall pay-offs.

(2) When playing with their partner, women are significantly more likely to give up their income autonomy and perform task B as opposed to the control group with

unfamiliar partners, where the majority chooses to perform task A. For men, no such differences are observed, which suggest that their behaviour is not affected by familiarity with their female partner. This is confirmed by comparing only cooperators in both groups: Familiar females' probability of performing the unpaid task is significantly larger than 0.5, whereas among unfamiliar female cooperators, no such deviance can be observed. Hence, unfamiliar cooperating women and men completed the unpaid task with equal probability, consistent with economic theories on household-specialisation decisions. Therefore, when the partners lack objective measures to detect relative productivity differences, only strangers divide tasks as predicted by the new home economics or cooperative bargaining models, whereas within couples, gender stereotyping seems to drive the decision.

(3) An analysis of income distribution over the two stages reveals the costs and benefits for realising efficiency gains through specialisation. The gendered pattern of labour division among familiar couples accounts for the gap that opens up in the partners' incomes at stage 1. The efficiency gains familiar couples realise by design come at the cost of financial autonomy of one of the partners, but the costs are not shared equally between men and women: Because female participants perform the unpaid task with a probability greater than one half, after the first stage, they receive (on average) a significantly lower income compared to familiar males, and also compared to unfamiliar females. These gaps close at the end of stage 2 because their partners generally behave reciprocally and 'reward' the assisting task-B-player by investing into the common pool. As a result, it is the familiar male who predominantly determines the familiar female's income; they act as "providers". Furthermore, they enjoy the benefits of specialisation and pooling: After stage 2, the income gap between familiar and unfamiliar men is the only one that retains significance and economic relevance.

The analysis of potentially confounding variables supports the robustness of the findings. No significant differences could be detected with respect to personality traits or attitude measures. Trust might provide a simple explanation for *why*

cooperation evolves more often among couples; not surprisingly, familiar participants report a significantly higher level of trust toward their partners. However, female trust levels do not differ from men's, neither in the familiar nor in the unfamiliar group. So if trust is the driving force behind cooperation, it is still unclear *how* it is driving which form of labour division couples choose, i.e., the gendered patterns in specialisation. Anticipation of gender-specific productivity differences should not drive the results, as an analysis of various personality traits and attitude measures has revealed that familiar females do not differ systematically from unfamiliar females. Therefore, even if women have biased beliefs about the productivity distribution by gender (i.e., perceive the male as more able), there is no indication that familiar females' beliefs exert a stronger bias in this respect than unfamiliar females' do. Thus, they should not have more reason than unfamiliar females to assume productivity differences relative to their partner.

Overall, the results point to some blind spots in conventional economic theory and the explanations it provides for gender gaps in various individual labour market outcomes. Theories that relate females' reduced level of participation to a) women's (anticipation of a) lower return for their participation compared to men or b) to a self-selection because of "female" preferences for non-market work (or a combination of the two), do not provide a sufficient explanation for the differences in behaviour between familiar and unfamiliar females in this experiment.

It seems plausible to consider social norms and the corresponding gender stereotypes to play a major, presumably subconscious role in driving participants' behaviour. Whenever framed within a social context that activates gender stereotypes (and heterosexual relationships might reasonably be claimed to create such a context), females—and eventually males, too—may be inclined to adjust their decisions in order to comply with the ruling social norms.[37]

[37]This is even more so the case, when individuals do not have access to complete information and potential outcomes of a decision are not entirely foreseeable. In the real world, the costs of deviating from social norms, and thus the benefits of complying, may weigh in immediately, whereas the costs for conform behaviour (in this case, women giving up financial autonomy when a man acts as provider) often occur in the future; employability decreases with each year spent

The results presented appear to be driven by social gender norms regarding the appropriateness of specialisation in one of the two tasks for the different sexes, thus entering an economic decision making process. This suggests that, even when expected returns from labour market activity are distributed equally between men and women, the probability to specialize in unpaid household-related labour might not be. With respect to real-world decisions, the findings indicate that social norms and gender stereotypes could account for the gender gap in family-work and labour market participation. This has direct policy implications, as one might argue that unequal labour division will continue to hinder female labour market success, despite increasing female educational success (for example), unless the potential efficiency gains that intra-household labour division promises become sufficiently small. Examples of how this could be achieved include, on the one hand, the abolition of policy instruments that encourage a breadwinner model explicitly (e.g., through differential taxation of spouses' incomes or forms of direct monetary incentives to substitute family-related market-services with home-production). On the other hand, it also requires the availability of affordable household-related services, such as childcare, to ensure that home production and labour division within households become less profitable.

outside the labour market, and labour market absence is associated with less accumulated savings and social insurance entitlements and, consequently, a higher risk for poverty.

3.9 Appendix A: Experimental instructions

You are playing this game with your partner. There are two stages to this game in which you can both make different choices.

Note: Your show-up fee of £4 will stay completely unaffected and will be paid out regardless the choices you make. The following instructions only refer to the earnings you can make on top of that.

Stage 1

You and your partner can both choose between two different tasks.

- Task A: A quiz, pays off 30p per correct answer.
- Task B: Assisting to task A, will be performed afterwards. It does not yield a pay-off in itself, but it increases the pay-off for the task A to 90p per correct answer.

You can either both choose to do task A independently or one of you can do task A while the other one does the assisting task B.

Examples:

(1) Assuming both of you do task A, if one of you answers 10 questions correctly, he/she will be paid out £3. If the other one answers 11 questions correctly, he/she will be paid out £3,30.

(2) Assuming one of you does task A and the other one performs the assisting task B afterwards, if the one who does tasks A answers 10 questions correctly, he/she will be paid £9. The other one will receive nothing...

Important note: You will receive your pay-off in private. Your partner will not get to know how many questions you answered correctly, regardless of which task he/she performs. This means he/she does not know how much you earned at this

stage and will also not find out later on.

Stage 2

In case you received a pay-off in stage 1, you can now decide how much of it you want to invest into a common pool. You can choose any sum between nothing and everything you received at stage 1. The amount invested into the pool will be increased by 20% and hereafter be equally distributed between the two of you.

Examples:

(1) One of you has done task A, the other one task B. Only the one who has done A receives a payoff, say £9. If you invest all of it into the common pool, this sum will be increased by 20%. Now there is £10.80 in the pool which will be distributed equally between the two of you, so that each receives £5.40.

(2) Both of you have done task A. Let's assume both of you receive £3. One of you invests all of it while the other one decides to invest £2. There are now £5 in the pool that will be increased by 20% to £6. Split in half, each of you receives £3. The one of you that invested all your pay-off hence has made £3 pounds in total, while the other one who kept £1 now has £4.

Testing your understanding:

(i) Assume both you and your partner do task A. You answer 5 questions correctly. How much money do you receive after stage 1 is completed?

(ii) Assume you do task A and your partner does task B. You answer 9 questions correctly. How much money do you receive after stage 1 is completed?

(iii) Assume you do task B and your partner does task A. Your partner answers 11 questions correctly. How much money do you receive?

(iv) Assume you have been paid out £5 after stage 1. If you invest all the money in the common pool and your partner invests £5, too, how much money will you be paid out after stage 2 is completed?

(v) Assume you have been paid out £5 and your partner did not get any pay-off because he performed the assisting task B. Assume you invest all your money in the common pool, how much will you be paid out after stage 2 is completed?

3.10 Appendix B: Psychological questionnaire items

To what extent do you agree with the following statements?

1. I am confident I get the success I deserve in life.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

2. Sometimes I feel depressed.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

3. Sometimes when I fail I feel worthless.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

4. I am filled with doubts about my competence.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

5. I determine what will happen in my life.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

6. Becoming a success is a matter of hard work; luck has little or nothing to do with it.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

7. In the long run, people get the respect they deserve in this world.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

8. When I make plans, I am almost certain I can make them work.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

9. What happens to me is of my own doing In my case, getting what I want has little to do with luck.

   ☐ I strongly agree.

   ☐ I rather agree.

   ☐ I rather disagree.

   ☐ I strongly disagree.

10. Many of the unhappy things in people's lives are partly due to bad luck.

    ☐ I strongly agree.

    ☐ I rather agree.

    ☐ I rather disagree.

    ☐ I strongly disagree.

11. Without the right breaks, one cannot be a good leader.

    ☐ I strongly agree.

    ☐ I rather agree.

    ☐ I rather disagree.

    ☐ I strongly disagree.

12. Who gets promoted often depends on who was lucky enough to be in the right place first.

    ☐ I strongly agree.

    ☐ I rather agree.

    ☐ I rather disagree.

    ☐ I strongly disagree.

13. Most people do not realise the extent to which their lives are controlled by accidental happenings.

    ☐ I strongly agree.

    ☐ I rather agree.

    ☐ I rather disagree.

    ☐ I strongly disagree.

14. Many times I feel that I have little influence on the things that happen to me.

    ☐ I strongly agree.

    ☐ I rather agree.

    ☐ I rather disagree.

☐ I strongly disagree.

15. How important is the chance you have to accomplish something worthwhile?

    ☐ Very important.

    ☐ Rather important.

    ☐ Rather unimportant.

    ☐ Not important at all.

16. How important is your chance at getting a promotion or getting a better job?

    ☐ Very important.

    ☐ Rather important.

    ☐ Rather unimportant.

    ☐ Not important at all.

17. How important is the friendliness of the people you work with?

    ☐ Very important.

    ☐ Rather important.

    ☐ Rather unimportant.

    ☐ Not important at all.

18. How important is the respect you receive from the people you work with?

    ☐ Very important.

    ☐ Rather important.

    ☐ Rather unimportant.

    ☐ Not important at all.

Chapter 4

Wage earners, homemakers & gender norms: What drives couples' labour division choices?[1]

Abstract

Why do only few couples choose the female spouse as main provider of labour income? To assess potential explanations, I study 246 subjects, real heterosexual couples, playing a specialisation game in the lab. Suprisingly, women are less likely to be chosen as breadwinners than men. The result is mainly driven by gender differences in productivity, but male overconfidence and women's reluctance to assume sole responsibility for household income appear to add to the underrepresentation of women among breadwinners. While I find little evidence that gender identity affects labour division choices, men's and women's performance suggests identity concerns may impact labour supply at the intensive margin: Men exert significantly more effort when their partner was previously more productive.

[1]Valuable comments by Miriam Beblo, Berno Büchel, Elisabeth Bublitz, Shushanik Margaryan, Eva Markowsky, Gerd Mühlheußer, Kristin Paetz, Arne Pieters, Claudia Schwirplies, and participants of the 2016 IAREP/SABE meeting, the 2017 meetings of ESA, SEHO, WEAI and UHH internal seminars are gratefully acknowledged. I thank the UHH WiSo Graduate School and Miriam Beblo for providing funds in support of this project.

4.1 Introduction

Although the gender gap in labour force participation has narrowed considerably in industrialised countries over the course of the past century, recent years have brought about little progress (OECD, 2017b). One important obstacle to the ultimate convergence is the gendered division of labour. In most families,[2] women spend more time providing unpaid care services and less time performing paid work than their partners (European Commission, 2015b; OECD, 2017b), with considerable implications for various economic outcomes: gender inequalities in wages, savings, and pension claims start widening in childbearing years and continue to grow over the life cycle (OECD, 2017a; Blau and Kahn, 2017; Goldin, 2014; Goldin et al., 2017; Kleven et al., 2018; Lundborg et al., 2017). While labour division itself may raise productive efficiency within the family, it is unclear why only a very small proportion of households reverse traditional gender roles and choose a female breadwinner[3] (OECD, 2017b). In this chapter, I use a lab experiment to study three potential motives: compliance to gender norms, male overconfidence, and female reluctance to assume sole responsibility for family income.

The first—and until today, most—influential economic models of household production, emphasise productive efficiency as families' main motive. Household members maximise the gains from trade by specialising in wage income and household production according to comparative advantage (Becker, 1973; Gronau, 1973a,b). Notably, the models predict equal proportions of male- and female-

[2]This chapter focusses on heterosexual couples, and so the terms 'family,' 'couple,' 'spouses,' or 'partners' are used interchangeably, and refer to cohabiting mixed-sex couples, married or unmarried. While unmarried heterosexual couples tend to specialise less than married, same-sex couples show the lowest degree of specialisation (Black et al., 2007). One reason for this pattern may be that the gains from specialisation are higher in the presence of children, and they are more likely present in married or unmarried heterosexual couples. Second, gains from specialisation may not be realised because access to legal marriage, and thus to binding contracts, is often restricted for same-sex couples.

[3]Throughout the chapter, I use a modified definition of the term "breadwinner" that does not necessarily imply full specialisation (only one spouse working in the market), but merely that the breadwinner spouse spends more time in paid work than their partner.

breadwinner households, *unless* one assumes systematic, gender-related productivity differences. More recently, economists became interested in an alternative hypothesis, based on the notion that individuals sometimes face a trade-off between income maximisation and their concerns for self-image or identity (Akerlof and Kranton, 2000; Bénabou and Tirole, 2011). In the presence of such concerns, couples' are not only striving to maximise the gains from specialisation, but also to conform with social gender norms, e.g. "a husband should earn more than his wife." When it comes to choosing a labour division, these two goals are perfectly aligned in families where the male spouse holds the comparative advantage in market production, but pose a conflict when the female partner is relatively more productive in the market.

Empirically, it is well established that culture and social norms affect female labour supply (Alesina et al., 2013; Fernández et al., 2004; Fernandez and Fogli, 2009; Fernández, 2013; Fogli and Veldkamp, 2011; Fortin, 2005, 2015), but the mechanisms are less well understood. While most of this literature conceptualises women as single decision-makers, few attempts have been made to synthesize the empirical findings on gender norms and theoretical models of family decision-making.[4] One is provided by Akerlof and Kranton (2000) in their seminal paper on identity economics. They discuss an application of their theory to the eco-

[4]The study by Fernández et al. (2004) forms an exception; they propose a theoretical model in which men, when their mother was employed during their childhood, exhibit either a higher household productivity or a greater taste for working wives. In both cases, the labour supply of married women increases, either because the composition of married women changes or because their relative advantage in household production decreases. Also, a small number of theoretical contributions have combined familily economic models with gender norm elements. Lundberg and Pollak (1993) were the first to model a non-cooperative "separate spheres" equilibrium, in which spouses' contributions to different household public goods are determined by social norms, as the threat point for cooperative bargaining in marriage. Alesina et al. (2011) present a bargaining model with a divorce threat point to study male and female labour supply, and show that women's might be relatively more elastic when social norms elevate the male partner's bargaining power above the female's. Cudeville and Recoules (2015) study male and female labour supply in a conjugal contract model that includes individual concerns for conformity with endogenously arising social norms and show that this might explain inefficiently low rates of female breadwinner households.

nomics of the household, although not explicitly modeling the couple's decision within a specific family economic framework. Using US income data, they show that husbands' time contribution to housework is not proportionately related to their share of labour market hours, as one may expect in a standard economic model.[5] The authors argue that this is because husbands lose identity utility if they deviate from social prescriptions such as "men should not do women's work" or "men should earn more than their wives," and wives therefore compensate by providing more housework. The study of Bertrand et al. (2015) provides an extensive set of empirical tests of this idea; most relevant in the context of this study is their analysis of married women's labour supply. Studying US census data, the authors find that, consistent with identity theory, wives are more likely to reduce their hours of work or leave the labour force entirely as the probability that they out-earn their husband (determined by the market value of their characteristics such as occupation, experience, etc.) increases.[6]

The finding that women reduce their labour market activity as they become more likely to out-earn their husbands is startling and warrants further investigation. To advance our understanding of how gender norms play into couples' labour division decisions, this chapter begins by constructing a simple collective model (Chiappori, 1988, 1992) of household production in the spirit of Becker (1973) and Gronau (1973a,b) that accommodates partners' potential concerns for identity (Akerlof and Kranton, 2000). The model is useful for understanding the most pressing issues for the empirical identification of identity concerns. Since produc-

[5]Note that this observation is not necessarily at odds with standard economic models if we consider non-cooperative models that include leisure consumption as a third means of time use alongside market and home production. Beblo and Robledo (2008) present such a model in which individual time contributions to household production are public good investments. When husbands have a first-mover advantage (because they are often older than wives), they will spend less time in housework and enjoy more leisure time than their wives, even when spouses spend the same amount of time engaged in market work.

[6]Bertrand et al. (2015) also show that women's time spent in housework increases in their potential to out-earn their husbands. Furthermore, they find that marital satisfaction decreases and the risk of divorce increases.

tive efficiency is determined by relative rather than absolute productivity differences, increasing women's wage productivity, even beyond their husbands', will generally not alter *comparative* advantage if women's absolute advantage in home production is large. Consequently, the observed reductions in women's labour supplies may be driven by a standard income effect rather than identity concerns. If, additionally, there are gender differences in preferences for the amount of resources invested in household production, the model predicts an even stronger negative effect. This is due to the assumption that spouses bargain cooperatively: As a wife's wage-earning capacity increases relative to her husband's, her bargaining power increases, and the allocation of household resources will reflect female preferences more strongly (see Manser and Brown, 1980; McElroy and Horney, 1981; Chiappori, 1988). As a result, the household will invest more heavily in home-produced goods, using the female partner's labour as input.[7]

Although the analysis of the model suggests that *explicit* concerns about identity are not the only plausible reason for reductions in women's labour supply as their earnings capacity increases, it does not imply that gender norms are irrelevant to this phenomenon in more conventional models of family decision-making. In fact, their impact on labour division choices may be considerable, but the mechanisms at play would be distinctly different: in identity models, individuals are aware of social prescriptions and actively try to avoid violating them, while norms may operate more subtly in standard models of family decision-making, by nurturing gender differences in household productivity and/or preferences.[8] The experimental approach used in this chapter draws on these insights; it attempts to

[7] A similar conclusion is reached if time spent outside the market is (mostly) leisure time and thus privately consumed. If women have a stronger preference for time spent outside the market, systematic gender differences in household productivity are no longer a prerequisite for generating similar predictions. Basu (2006) argues that the balance of power both determines, and is in turn determined by, the woman's labour income, not just her earnings potential. This also implies that she can realise the same improvement in her bargaining position with fewer hours of work as her wage increases.

[8] Gender norms may impact women's productivity advantage in the home, e.g., by increasing the *perceived* productivity advantage women have in raising children. See Section 4.2 for details.

systematically isolate the identity channel by deliberately muting the influence of household productivity and preferences for time spent outside the market.[9] Thus, the first important question addressed is whether the identity hypothesis can be confirmed in such a setting. If so, a second important question is whether men and women are equally concerned about identity. While Akerlof and Kranton (2000) assert that it is the *husband's* identity utility that suffers if his wife out-earns him, the empirical analyses provided so far (Bertrand et al., 2015; Wieber and Holst, 2015; Hederos Eriksson and Stenberg, 2015; Görges, 2015) are able to study identity concerns only at the couple level. Opening this black box by studying the labour division decision—rather than simply labour supply—of men and women at the *individual* level therefore represents an important contribution of this chapter.

The experiment was conducted with 246 participants, i.e., 123 real heterosexual couples, who played a specialisation game that mimics the time allocation decision between market work and home production analysed in the model. Each individual played with (a) their partner, (b) a randomly assigned stranger of the same sex as the partner, and (c) a computer. The outcome I am interested in is spousal labour supply at the extensive and intensive margins. Regarding the former, I analyse labour division choices, which can be one of three types: (i) dual earner, where both partners supply labour to the market and earn income for the couple, (ii) female earner, where only the female partner supplies labour to the market, while the male partner engages in household production, or (iii) vice versa in the male earner option. Conditional on productivity differences existing within the couple,

[9]The experimental design has several advantages as it allows to rule out gender differences in preferences for performing labour market work versus housework, gender differences in productivities, and productivity-based selection into marriage. In the lab setting, gender-specific preferences for the tasks are precluded because the tasks performed in the lab are neutral, and this assertion can actually be checked in the data because participants are asked to rate the tasks after a trial round. Finally, selection into the couple in the real world may be endogenous to (expected) labour market and household productivities. While this holds true as well for couples that participate in the present study, the selection is plausibly independent of partners' respective productivities in the two lab tasks. Consequently, the lab experiment proposed in this chapter allows to cleanly identify the effect of a gender norm "a man should not act as a homemaker" on couples' specialisation decisions, above and beyond these alternative explanations.

by design, only one choice is income-maximising (female or male earner, depending on which partner is more productive in wage earning). Since income maximisation and identity goals are at odds only when the productivity advantage is held by the female partner and not when held by the male, comparing subjects' choices across these types of couples reveals whether identity concerns play a role. Only if this is so should the probability of choosing an income-maximising labour division be lower for subjects in female productivity advantage couples. The probability of choosing an income-maximising labour division should only be lower if identity concerns do influence the decisions made by these couples.

The results show that women are less likely to become breadwinners than men. Unpacking the mechanisms, I find that this is largely due to productivity differences in the market task; when presented with information about one's partner's productivity in a previous stage, subjects generally do not seem to treat a productivity advantage held by a man differently from one held by a woman. When full information is not available, however, all subjects make less efficient choices. While women are equally likely to make mistakes regardless of the gender of the partner holding a productivity advantage, male subjects are significantly more likely to make a mistake when it is held by the female partner. This pattern suggests that, to some extent, women holding a productivity advantage in the market are less likely to become breadwinners than men. However, the driver seems to be men's overconfidence rather than identity concerns. Thus, this chapter is the first study to establish that male overconfidence may not only contribute to explaining men's overrepresentation in certain top positions, as argued by Niederle and Vesterlund (2007), but also their overrepresentation among breadwinners.

Analysing choices with computer partners reveals a second surprising result: Women choose themselves as breadwinners at much higher rates when paired with a computer than when paired with real men. I present evidence that women might shy away from the single-earner role because they feel more pressure when performing as the sole provider, compared to merely co-contributing to the couples'

income as a dual earner. Finally, a complementary analysis of effort supply suggests that, while identity concerns may not impact the labour division choices, and thus not the labour supply of men and women at the extensive margin, they may affect the intensive margin. I find that men increase their effort supply substantially when their partner was previously more productive, i.e. when the threat of being out-earned is high. Since this effect is only present when they are paired with their real partner, and not with a stranger or a computer, it appears to reflect men's desire to conform with the norm "men should earn more than their wives."

The chapter is organised as follows: Section 4.2 outlines the theoretical framework of the chapter. The experimental design, recruitment and sample characterisitics are detailed in Section 4.3. Sections 4.4 contains the main results and further investigation of the mechanisms. Section 4.5 concludes.

4.2 Theoretical background

To conceptualise how identity concerns may influence couples' labour division decisions, I draw on a standard household production model (Becker, 1973; Gronau, 1973a,b), place it within a simple collective framework of family decision-making (Chiappori, 1988), and introduce an identity component in the spirit of Akerlof and Kranton (2000). In the model, spouses maximise their joint utility, which is an increasing function of household income and public good consumption, and, possibly, compliance with gender norms (i.e., identity). They decide collectively on the optimal allocation of each partner's time to wage income or household production, and on the division of the surplus. This theoretical framework, albeit very simple, allows for testable predictions regarding the impact of gender identity on couples' specialisation decisions. The model is also useful for exploring other behavioural motivations that may explain the persistence of traditional labour division choices. This is instructive for pinpointing the most pressing issues for the empirical identification of identity concerns and thus informs the design of the

experiment presented in Section 4.3. A number of factors are shown to give rise to traditional labour division, even absent any explicit concern for gender identity.

4.2.1 A collective model of household production with identity

Consider a household comprising two decision makers: male m and female f.[10] Spouses derive utility from private consumption of a Hicksian composite good, c_i, a primary household public good, G, and potentially from identity, I. Individual utilities can be represented by the function $U_i(c_i, G; I)$, $i \in \{m, f\}$. To focus the discussion on the effect of identity on labour division, I assume that c_i can only be purchased in the market, G can only be produced in the home, and, moreover, abstract from leisure. Thus, spouses can allocate time to two distinct activities: (i) paid market work that produces wage income and (ii) unpaid household work that produces the primary household public good. Although there are potentially many public goods in a household, it can be useful to think of G as childcare, where a greater provision of the public good requires a larger time investment and increases child well-being (Browning et al., 2014).[11] With regard to identity, I follow Akerlof and Kranton (2000) in assuming that agents derive utility from matching behavioural prescriptions. To keep the model tractable, I impose strict separability of utility, thus:

[10] Children may be present in the household but are not considered to actively influence time allocation decisions, or to contribute themselves to the household production process.

[11] Browning et al. (2014) argue that children are a useful example for a public good because their well-being arguably increases the utility of both parents, i.e., its 'consumption' is non-rival. In the case of divorce, both parents continue to derive utility from their children, even if the parent holding custody may do so to a greater extent. Here, I implicitly assume that time allocated to the production of the public good increases child well-being, and thus child *quality* in the sense of Becker and Lewis (1973), but higher time investments can also be interpreted as producing higher child *quantity*.

$$U_i(c_i, G) = u(c_i) + \delta_i v(G) + \psi_i I \tag{4.1}$$

The parameter $\delta_i > 0$ allows spouses to differ with respect to how much they value the public good. Similarly, $\psi_i \geq 0$ is a constant that represents the weight the individual places on identity, and may also differ between husband and wife. If an individual does not care about identity at all, i.e., $\psi_i = 0$, the utility function reduces to its standard components u and v. For both these functions $f = \{u, v\}$, I assume $f' > 0$, $f'' < 0$, and $f'(0) = \infty$, i.e., c_i and G are normal goods.

Total time endowment, normalised to 1, places a constraint on time allocations, i.e., producing childcare will naturally come at the cost of forgone labour earnings and vice versa. The level of private consumption is determined by the market price for the composite good, c_i (normalised to one), the individual's non-labour income, n_i, her wage rate, w_i, and time spent working in the labour market, t_i:

$$c_i = w_i t_i + n_i \tag{4.2}$$

The level of home-produced childcare, G, is determined by the spouses' time inputs, $1 - t_i$, and productivities, h_i. Time inputs of m and f are assumed to be perfect substitutes[12], thus:

$$G = h_m(1 - t_m) + h_f(1 - t_f) \tag{4.3}$$

[12]This assumption is very common in the literature, even though complementarities may exist, particularly in parenting (Browning et al., 2014). Constant returns to scale in production is a necessary assumption to reproduce the key properties of household models à la Becker (1973), see Pollak and Wachter (1975) and Apps and Rees (2009).

Finally, I follow Akerlof and Kranton (2000) and Bertrand et al. (2015) in assuming that the level of identity utility is determined by compliance with the norm "a husband should earn more than his wife." Specifically, I model (non-)compliance with this norm as imposing a psychic reward (cost),[13] i.e., the identity term becomes negative (positive) when the husband's earned labour income is smaller (greater) than the wife's. For simplicity, I assume that the total psychic costs (rewards) increase linearly in the distance between the labour incomes of m and f:

$$I(t_m, t_f, w_m, w_f) = t_m w_m - t_f w_f \tag{4.4}$$

Moreover, I assume that this identity component is not present in the utility functions of singles, and thus does not affect spouses' outside option. Note that "producing identity," unlike producing wage income or the primary household public good, is not an activity in its own right, but rather a by-product of spouses' labour division. Furthermore, notice that, in the context of household production, identity has the features of a public good, as its consumption is non-rival and non-exclusive.[14]

Given the assumptions made regarding their individual utilities, how do m and f determine individual time allocations to market and home production?[15] In col-

[13]The approach is similiar to the one used by Akerlof and Kranton (2002) in their identity model of schooling decisions.

[14]Browning et al. (2014) discuss the "subtle interactions between the ('technical') nature of a good and how it enters (...) utilities" (81) and note that one partner's consumption of a private good may have a negative externality on the other, while some goods that have public good character in principle are truely only consumed by one partner. This applies to identity concerns, as underlined by the discussion Akerlof and Kranton (2000) provide regarding the negative externalities a norm violation of i may cause for the utility of j.

[15]Naturally, the answer to this question hinges on the spouses' ability to cooperate, but non-cooperative models of spousal time allocation also offer explanations for why a wife might produce more childcare and work less in the labour market than her husband. For example, Lundberg and Pollak (1993) assert that non-cooperative spouses choose their contributions to specific public goods according to gender norms. Others assume spouses choose their individual

lective models, the division of the jointly produced surplus typically cannot make a partner worse off than his or her best alternative; i.e., each spouse is assumed to maximise her utility subject to a participation constraint of her partner. Browning et al. (2014) emphasise that, in doing so, collective models rely on fewer assumptions regarding the precise nature of the bargaining process than Nash bargaining models (McElroy and Horney, 1981; Manser and Brown, 1980). The latter class of models assumes that spouses maximise the Nash product, i.e., the product of the individual differences between their consumption level within and outside the marriage. The former typically just assume that there is *some* stable sharing rule for the division of the joint surplus, but it can be equivalently obtained by maximising the Nash product (Chiappori, 1988). The set of solutions to this program includes all points on the Pareto frontier; the partners' respective Pareto weight, the so-called distribution factors, determines which specific point will be selected (Browning et al., 2014). The distribution factors reflect individual opportunities outside marriage, which affect distribution therein. It has been shown that this program is technically equivalent to maximising a household welfare function, i.e., the sum of spouses' utilities, weighted by the distribution factors (Chiappori, 1988; Blundell et al., 2005). Denoting the wife's distribution factor by α while normalising the husband's to 1, the individual utility functions can be combined into a household welfare function[16]:

utility functions under the Nash conjecture, either entirely selfishly (Konrad and Lommerud, 1995), or under mutual caring (Chen and Woolley, 2001). Formally, individual i chooses her optimal time investment $1 - t_i$ into G, taking her partner's investment $1 - t_j$ as a given. While non-cooperative models generally predict underprovision of the public good, we may still observe time investments that resemble specialisation. This may be due to gender norms as in Lundberg and Pollak (1993) or to price differences as in Konrad and Lommerud (1995). The latter show that spouses' contributions are not only determined by relative productivities, but also by absolute advantages.

[16]Again, I abstract from the marriage market effects here and follow Engineer and Welling (1999) in assuming that "love strikes randomly", and that marriage allows partners to generate higher utility than staying single.

$$W = U_m(c_m, G; I) + \alpha(w_f, n_f, w_m, n_m, \boldsymbol{z})U_f(c_f, G; I) \tag{4.5}$$

The definition of α as a function of wages, non-market income, and other factors that influence intra-household bargaining power, captured in the vector $\boldsymbol{z}$,[17] is crucial to a collective model. Notably, certain factors that typically affect household demand in a unitary setting,[18] such as prices (including wages), and non-labour income, also affect spouses' outside options (i.e., participation constraints), and thus the relative weight placed on the wife's utility. Consequently, household demand may respond differentially to, e.g., an increase in non-labour income accruing to the husband versus the wife, if the spouses have different preferences over consumption.

Because the nature of the bargaining process is cooperative, monetary transfers $\tau \gtrless 0$ between spouses are possible. This guarantees that the level of the public good G will be efficient given spouses' weighted preferences. It will also be *produced* efficiently, i.e., by the spouse who imposes lower opportunity costs in terms of forgone wage earnings on the household. To see this, we can use the fact that

[17]These include, among others, divorce and alimony laws and spouses' marriage market opportunities, as well as social norms regarding "say" in a marriage (Basu, 2006).

[18]The first economic models of family decision making regarded the family as an economic unit, and either assumed that spouses have identical preferences (Samuelson, 1956), i.e., no conflict of interest, or that α is fixed because it is simply decided upon by the benevolent household dictator as in Becker (1974b).

$c_f = w_f t_f + n_f + \tau$ implies $\tau = c_f - w_f t_f - n_f$,[19] and equations (4.1)–(4.5) to obtain the couples' maximisation problem:

$$\begin{aligned}\max_{t_i \in [0,1]} W = &\Bigg[u\Big(w_m t_m + n_m - c_f + w_f t_f + n_f\Big) + \\ &\delta_m v\Big((1-t_m)h_m + (1-t_f)h_f\Big) + \psi_m\Big(t_m w_m - t_f w_f\Big)\Bigg] + \\ &\alpha(\cdot)\Bigg[u\Big(c_f\Big) + \\ &\delta_f v\Big((1-t_m)h_m + (1-t_f)h_f\Big) + \psi_f\Big(t_m w_m - t_f w_f\Big)\Bigg]\end{aligned} \tag{4.6}$$

The Kuhn-Tucker conditions for the t_i give the optimal time each spouse devotes to market production (to focus the discussion on male breadwinner cases where f spends at most as much time in the labour market as does m, conditions for $t_f > t_m$ are suppressed):

$$\begin{aligned}\frac{\partial W}{\partial t_f} =& u'(c_m)w_f - (\delta_m + \alpha(\cdot)\delta_f)v'(G)h_f - (\psi_m + \alpha(\cdot)\psi_f)w_f \leq 0 \\ 0 \leq t_f <& 1 \\ t_f \frac{\partial W}{\partial t_f} =& 0\end{aligned} \tag{4.7}$$

[19] Note that the optimal transfer τ will equalise spouses' marginal utility from private consumption according to the weights of their utility functions: $\alpha(\cdot)\frac{\partial u(c_f)}{\partial \tau} = \frac{\partial u(c_m)}{\partial \tau}$. Thus, the wife's share of monetary consumption increases in α, and whenever $\alpha = 1$, spouses will share money income for private consumption equally.

$$\begin{aligned} \frac{\partial W}{\partial t_m} =& u'(c_m)w_m - (\delta_m + \alpha(\cdot)\delta_f)v'(G)h_m + (\psi_m + \alpha(\cdot)\psi_f)w_m \geq 0 \\ 0 < t_m \leq& 1 \\ (1 - t_m)\frac{\partial W}{\partial t_m} =& 0 \end{aligned} \tag{4.8}$$

Because $G'(0) = \infty$, the public good will always be produced. Notice that identity utility increases in male and decreases in female market time, while both have a positive effect on private consumption utility and a negative effect on public good consumption utility. Four types of male breadwinner labour division choices are possible (i.e., m spending weakly more time in labour market production than f):

(i) Both spouses produce G, only the husband also works in the labour market: $t_f = 0$, $t_m \in (0, 1)$.

(ii) The husband only works in the market, the wife only produces G: $t_m = 1$, $t_f = 0$.

(iii) Both spouses work in the market, only the wife also produces G: $t_m = 1$, $t_f \in (0, 1)$.

(iv) Both spouses work in the market *and* produce G: $t_i \in (0, 1)$, for $i = m, f$.

Using conditions 4.7 and 4.8, Table 4.1 presents an overview of the effect of an increase in the wife's wage rate on time allocations with and without identity concerns. For illustration purposes, it suffices to study one out of the four types of male breadwinner labour division: the case of partial specialisation (iii). Because identity concerns are driven by her potential to out-earn him, and may thus affect her relative contribution to household labour income, the table also includes two indicators for relative income. Let $S_f = \frac{w_f}{w_f + w_m}$ denote the wife's potential relative

contribution to "full" household labour income (in the sense of Becker, 1965), and $\tilde{S}_f = \frac{w_f t_f}{w_f t_f + w_m t_m}$ be her contribution to realised household labour income. By definition, S_f always increases as her wage rate increases, but $\tilde{S}_f$ depends also on the change in her labour supply t_f (and potentially his), which in turn depends on the parameters of the model.

Table 4.1: Effects of increases in wives' wage rate on her time in home production and contribution to total household income in the partial specialisation case (iii)

condition	identity	valuation of G	$\frac{\partial t_f}{\partial w_f}$	$\frac{\partial t_m}{\partial w_f}$	$\frac{\partial \tilde{S}_f}{\partial w_f}$	$\frac{\partial S_f}{\partial w_f}$
$\frac{w_f}{h_f} < \frac{w_m}{h_m}$	$\psi_i = 0$	$\delta_f = \delta_m$	$\lesseqgtr 0$	$= 0$	> 0	> 0
	$\psi_i = 0$	$\delta_f > \delta_m$	$\downarrow$	$\rightarrow$	$\downarrow$	$\rightarrow$
	$\psi_i = 0$	$\delta_f < \delta_m$	$\uparrow$	$\rightarrow$	$\uparrow$	$\rightarrow$
$\frac{w_f}{h_f} \leq \frac{w_m}{h_m} \frac{u' + \psi_m + \alpha\psi_f}{u' - \psi_m - \alpha\psi_f}$	$\psi_f = \psi_m$	$\delta_i \approx 0$	$\lesseqgtr 0$	$= 0$	$\lesseqgtr 0$	> 0
	$\psi_f > \psi_m$	$\delta_i \approx 0$	$\downarrow$	$\rightarrow$	$\downarrow$	$\rightarrow$
	$\psi_f < \psi_m$	$\delta_i \approx 0$	$\uparrow$	$\rightarrow$	$\uparrow$	$\rightarrow$

Note: The last four columns display the partial derivatives of the wife's (t_f) and husband's (t_m) time in the market, her contribution to full ($\tilde{S}_f$) and to realized (S_f) household labour income, all with respect to her market productivity (w_f). In both the upper and the lower panel, the first row indicates the sign of the partial for the benchmark case reported in that row, while the arrows in the second and third row indicate how the partial changes relative to the benchmark.

The upper panel of Table 4.1 refers to the case without identity concerns, i.e., $\psi_i = 0$, $\forall i \in \{m, f\}$. Here, the condition for a male breadwinner labour division to be chosen is that the wife's relative market productivity is strictly smaller than

the husband's, i.e., $\frac{w_f}{h_f} < \frac{w_m}{h_m}$.[20] The first row considers the case where spouses have equal preferences for the public good, i.e., $\delta_f = \delta_m$. Because the wife spends some of her time working in the market, an increase in her wage rate triggers an income and a substitution effect; the production of G becomes relatively more expensive, but the household can also afford more of it. Therefore, the labour supply of f may increase or decrease, as shown in column $\frac{\partial t_f}{\partial w_f}$. The next column, $\frac{\partial t_m}{\partial w_f}$, shows that the husband's labour supply is unaffected. This is because the couple pays a higher price per unit of G when it is produced by the husband as compared to the wife and, because $1-t_f < 1$, any increase in demand for G will be met by the wife and not the husband (Gronau, 1973b). The effect on her realised contribution to household labour income will be positive, given the assumption that both c_i and G are normal goods. The second and third rows show how these effects change if there is a conflict of interest between the spouses, i.e., when the shift in bargaining power ($\alpha \uparrow$) triggered by the increase in her wage rate becomes relevant. If she has a stronger preference for G, i.e., $\delta_f > \delta_m$, demand will go up more strongly relative to the benchmark case of equal preferences, as will her time spent in its production. Consequently, her labour supply will be lower than in the benchmark, and, by the same logic, her realised contribution to household labour income. The husband's labour supply remains unaffected, for the same reason as before. Finally, if she has lower preferences for G than him, household demand will reduce and she will spend relatively more time in the market.

The lower panel of Table 4.1 considers the case where identity concerns are present. Here, men can be relatively less effective at wage earning and still spend *more* time in the labour market than their spouse. This can be seen from the condition for spouses to choose labour division (iii) in the presence of identity, and noting that $\frac{u'+\psi_m+\alpha\psi_f}{u'-\psi_m-\alpha\psi_f} > 1$. For convenience, suppose that spouses have a very low, identical valuation of G; thus, any time that f spends in home production

[20]This holds true for the cases (i) and (ii) as well. Note that it follows from the Kuhn-Tucker conditions (4.7) and (4.8), that only weak inequality is required, but the special case of equal productivities will be discussed separately below.

can be attributed to identity concerns. The first row of this panel again considers the case where no conflict of interest exists between spouses. As before, the effect of an increase in w_f on t_f and $\tilde{S}_f$ is ambiguous. The wife will reduce her labour supply if the marginal reduction in identity utility caused by her wage increase, which equals ψ, outweighs the marginal increase in consumption utilities. As is the case without identity, the husband's labour supply remains unaffected, as a further increase to make up for the identity loss is not feasible. A conflict of interest among spouses also has similar effects because, again, her wage increase will shift bargaining power in her favour. If the female partner's identity preferences exceed the male's, the household will increase its "demand" for identity and lower her labour supply as well as her realised contribution to household income $\tilde{S}_f$ relative to the benchmark case with identical identity preferences. Finally, if her valuation of identity is lower than his, demand for identity decreases and the effects go in the opposite direction.[21]

The analysis shows that an increase in the wife's wage rate relative to her husband's can have negative effects on the wife's labour supply, even without explicit concerns for identity. We saw that such a reduction is possible whenever the household has a relatively strong preference for the public good, and all the more plausible if her valuation exceeds his. Predictions are unambiguous *only* in the special case where $\frac{w_f}{h_f} = \frac{w_m}{h_m}$, as only the presence of identity concerns may then lead to a reduction in her labour supply. To see why, note that any labour division that produces the optimal amount of G is feasible in the case of identical productivities and no identity concerns, including the one considered in Table 4.1.[22] An increase in her wage rate changes the productivity ratios; she now produces

[21] If the valuation for the public good were higher, it is also possible that his labour supply reduces if she has a lower taste for identity and higher taste for G.

[22] The equal labour market participation case (iv) is a special case that spouses choose if and only if there are *no* differences in relative productivity. While this is a necessary, it is not a sufficient condition, as $t_m = t_f$ is not a unique optimum in the case of identical productivity ratios. The optimal provision of the public good, for given preferences, wages and household productivities, requires a total of time inputs $\tilde{t} = \tilde{t}_m + \tilde{t}_f$. As a consequence, any other combination of t_m and t_f that adds up to $\tilde{t}$ is also feasible, including (iii).

wage income relatively cheaper than her husband. Therefore, her labour supply will increase while his will decrease, regardless of whether there is a conflict of interest between them. As noted before, the absence of productivity differences constitutes a special case, and thus testing for identity preferences empirically is difficult without specific knowledge of spouses' productivities in the market and in the home, or of their preferences for the home-produced good. The problem is aggravated if such differences are themselves affected by social norms, as discussed in the following subsections.

4.2.2 The role of gender differences in preferences and productivity

To emphasise the problem that arise for the empirical identification of identity concerns, this subsection briefly discusses how gender norms may affect gender differences in preferences for the home-produced good, as well as in market and home productivites.

Preferences

The main conclusion from the model discussed above is that identity concerns may not necessarily drive the observation that women decrease their labour supply as their potential to out-earn their husband increases. A reduction was shown to be all the more likely when she exhibits a relatively stronger preference for the public good, G, than him. Given that G was defined as childcare, this seems to be backed by empirical evidence, as a number of studies has in fact shown that households tend to allocate more resources to child well-being as women's bargaining power improves (Lundberg et al., 1997; Phipps and Burton, 1998; Attanasio and Lechene, 2002; Duflo, 2003; Quisumbing and Maluccio, 2003). Still, the conclusion that the reduction in women's labour supply is not necessarily caused by identity hinges on the assumption that the wife is relatively more effective in the production of G, but

it is easy to modify the model such that productivity differences are not a premise. For instance, suppose that utility is not only derived from the public good once it is produced (i.e., a well cared for child), but rather from engaging in the productive process itself. In the extreme case, we may assume that spending time engaged in childcare is a pure leisure activity, and as such, time spent outside the market generates utility for the agent, just as with a private consumption good.[23] If she has a strong preference for spending time in childcare relative to market work, she may decrease her labour supply. Women may derive greater utility from spending time in childcare either due to biological gender differences[24], or as a result of social norms (Borck, 2014). Chapter 1 has shown that the gender gap in preferences for producing wage income relative to investing in children and family can indeed be influenced by political regimes and social norms. It is important to note that, although it is likely that many people derive some pleasure from time spent caring for their own children, the same may apply to time spent engaged in paid work. Consequently, these aspects are confounders of any labour division decisions that are hard to control for using micro-data. Studying labour division in a context where both activities are neutral in the extent to which they produce pleasure is a unique opportunity offered by a controlled lab experiment. This allows us to focus on the productive aspects of both activities; accordingly, the causes of gender differences in productivities are discussed in the following subsections.

[23]This is essentially the original collective model: Chiappori (1988) has shown that spouses maximise their individual utility subject to their individual budget constraint, which in turn is derived from splitting total non-labour income collectively, i.e., according to the sharing rule. In this version of the model, an increase of the wife's wage rate raises her Pareto weight, thereby increasing her share of non-labour income. Additionally, the standard income and substitution effects are triggered by the wage-related expansion of her budget set.

[24]Evolutionary biologists have long argued that men face greater uncertainty than women regarding their reproductive success, and thus have a lower willingness to invest in any given child (Trivers, 1972).

Household productivity

Whether or not systematic (and meaningful) gender differences in household productivity exist is disputed. Becker (1991) has argued that women's home productivity advantage derives from their reproductive abilities, i.e., the bearing and rearing of children, and can thus hardly be matched by men. On the other hand, Greenwood et al. (2005), Greenwood et al. (2005), and Albanesi and Olivetti (2016) present evidence suggesting that technological and medical progress have considerably reduced women's productivity advantage in the home. While this is certainly true, empirical evidence for *equal* productivity in the home is hard to come by, as it is almost impossible to measure. The discussion is further complicated by the fact that social gender norms and women's household productivity are likely to be correlated on three different levels: *perceived* and *de facto* gender differences in ability, and the availability of market substitutes. For example, the fact that the findings provided by Bertrand et al. (2015) were replicated for West Germany, but neither for East Germany (Wieber and Holst, 2015) nor Sweden (Hederos Eriksson and Stenberg, 2015), might be interpreted as evidence for the identity mechanism, which appears to not be at work in countries where gender norms are more progressive. At the same time, the availability of market substitutes for women's time spent in childcare is higher in Sweden and in East Germany compared to West Germany, or the US (Aisenbrey et al., 2009). This presents an alternative explanation for the different findings, since wives' relative productivity advantage in the home might be less relevant for specialisation if they are outperformed by the market (Gronau, 1973b).

Second, social norms may increase women's *perceived* productivity advantage in the home by asserting that the mother's time input is essential to the production of child well-being. In the context of the model discussed here, this could mean that women's expected returns to time spent in home production are extremely high, even when female and male labour market productivities are equal, which makes it plausible that the income effect outweighs the substitution effect if her

wage rate increases. Studies show that beliefs about the importance of maternal time as input in the production of child well-being varies greatly across cohorts and countries (Treas and Widmer, 2000; Bauernschuster and Rainer, 2012; Fortin, 2015) and this likely affects maternal labour supply (Fogli and Veldkamp, 2011; Fernández, 2013; Borck, 2014). Finally, if gender norms lead parents to train female children more intensively in housework and care-taking than male children, this may amplify *de facto* gender differences in home productivity (Hadfield, 1999; Engineer and Welling, 1999). Empirical evidence seems to support this; e.g., Solaz and Wolff (2015) find that, in a sample of French adolescents co-residing with their parents, girls on average spend one hour per day more on housework than boys. In light of these considerations, the experiment discussed in this chapter presents a unique opportunity to study labour division in a setting where productivity measures for home production are available, and, by design, irrelevant for maximising productive efficiency.

Market productivity

Consider two spouses of identical labour market "ability".[25] In the model presented above, labour income productivity is measured by individual wage rates, w_i.[26] One

[25]In this subsection, I abstract from the effect of selection into marriage on intra-couple productivity differences and follow Engineer and Welling (1999) in assuming that "love strikes randomly." It is, however, important to note that women of high labour market productivity are less likely to marry—and that this effect also appears to be amplified in countries where gender norms are more traditional (Bertrand et al., 2018). Moreover, husbands tend to be older than wives (Lundberg et al., 2016; Beblo and Robledo, 2008), which means their wages are higher on average due to more labour market experience and/or on-the-job-training.

[26]Using wage as an indicator for productivity is of course in itself problematic, since wage data are often endogenous to the specialisation decision. The two are often measured at the same point in time, and even with panel data it is hard to identify when, and based on which initial measures of productivities, a decision on specialisation was made. Using predicted wages based on demographic characteristics, as done by Bertrand et al. (2015),may alleviate the problem; however, the decisions underlying these characteristics (e.g., how much human capital has been acquired by each spouse) may have also been based on preferences for the public good and expectations about future specialisation. Many studies show that women choose lower-paying occupations and fields (Fortin, 2008; Grove et al., 2011; Humlum et al., 2012), and one of their motives may be anticipation of the time they will spend in household production.

important demand-side factor that causes wages to differ despite equal ability is labour market discrimination. While some studies suggests that it may not (any longer) be a primary driver for the contemporary gender wage gap in the US (Goldin, 2014; Cook et al., 2018), discrimination may in fact be larger in countries where more traditional gender norms prevail (Adams et al., 2017). This might be because gender norms inflate employers' "taste" for discrimination (Becker, 1957), or because they increase employers' incentives to discriminate statistically (Arrow, 1972; Phelps, 1972). Lazear and Rosen (1990) show that a gender earnings differential given equal ability in the labour market arises when women are more productive in the household. Because employers expect them to leave their job at a higher rate than men, more able women are less likely to receive promotions than less able men. As discussed above, the expected value of women's time in home production may be positively correlated with traditional gender norms, and so might employers' incentives for statistical discrimination.

Supply-side factors may also affect gender differences in (perceived) labour market productivity, e.g. because women choose to work in different jobs than men. Over the past few decades, much economic research on gender differences in preferences has been carried out in the lab and women's behaviour has been observed as, on average, less confident, risk-taking, and competitive than men in a variety of contexts (for a comprehensive overview, see Niederle, 2016, but note also the illuminating disscussion of the question of whether these differences are economically large and meaningful provided by Nelson, 2015). It has been argued that these differences may exacerbate the gender gap in labour market outcomes (Bertrand, 2011; Blau and Kahn, 2013), particularly with regard to the types of jobs men and women of *equal ability* choose (Reuben et al., 2017; Wiswall and Zafar, 2017). Notably, as was the case for the demand-side factors, these differences appear to widen in societies or contexts where more traditional gender norms prevail (see, e.g. Gneezy et al., 2009).

Apart from the most obvious reason for market productivity differences to arise, i.e., equally able wives choose to work in less lucrative jobs than their husbands, one finding from the experimental literature stands out as particularly interesting in the context of the collective model of labour division presented above: gender differences in over-confidence (Niederle and Vesterlund, 2007). Consider, e.g., a labour market in which principals determine wages using tournaments (Lazear and Rosen, 1981). When husband and wife are equally able and thus have the same expected earnings, men might be overly optimistic about their rank in the performance distribution, and thus consider themselves more likely to win the tournament than their wives. This way, differences in expected labour market productivities may result in traditional labour division, but based entirely on gender differences in confidence. Empirically, this seems to be supported by studies that show, using survey data, lower expected future pay in female graduates relative to their male counterparts, conditional on majors and grades (Reuben et al., 2017).

In its simplest version, the model presented above cannot readily capture the impact of gender differences in preferences for risk or competition on labour division. This would require allowing for a "psychological" cost or reward from one's *own* labour market activity, which may differ between genders because of, e.g., men's greater taste for competition, relative to women.[27] In a similiar vein, it is possible that women experience additional psychological costs from working in the

[27]To see why this is the case, consider the example of risk-aversion: In the collective model framework analysed here, the returns to labour market production may involve uncertainty, while the returns to home-production are certain, as in the model presented by (Vesterlund, 1997). The source of uncertainty in wage production could stem from the way time spent in wage labour is mapped into earnings, e.g., if firms reward employees according to output, and the level depends on the amount of time an agent invests given her productivity and some stochastic component. So long as the random productivity shocks follow the same distribution for men and women, gender differences in risk aversion will *not* affect labour division in a collective model. A higher risk aversion among women will *only* affect their preferences for total household time invested in the risky wage income production versus total time spent in home production, thereby affecting the level of the household demand for the home-produced good. Risk-aversion is, however, irrelevant for the decision which spouse works in the market for whatever time the household optimally allocates to wage production, which depends on the productivity ratios only. Again,

labour market if they are responsible for producing the lion's share of the family's wage income. Anecdotal evidence suggests that some women may perceive a higher level of stress and pressure in the role of breadwinners, a notion that has not yet received much attention from economists and one that this experiment was designed to study.[28]

Since all of the factors discussed above affect labour division choices by creating real or perceived gender differences in labour market productivity, but are unrelated to identity concerns, the experiment presented in this chapter is designed such that demand-side as well as certain supply-side factors are muted. The design allows for the intentional study of the impact of two specific factors on labour division decision: gender differences in over-confidence and women's psychological costs from assuming (sole) responsibility for family earnings. Section 4.3 provides details about the experimental design and procedures.

4.3 The experiment

4.3.1 The game

The basic structure of the game simulates the household production process as follows: Two players, m and f, can perform one of two activities, market work

this conclusion depends on the assumption that spouses fully cooperate, and is unlikely to hold in a noncooperative framework.

[28]The writer Ashley C. Ford wrote a piece entitled "Millennial Women Are Conflicted About Being Breadwinners" for the digital media companey refinery29.com in 2018, for which she conducted an anonymous survey with 130 women who out-earned their partners. A recurring theme among these women associated with breadwinning appears to be stress and pressure. One woman is quoted saying "It's stressful. It's a huge responsibility. I pressure myself to stay in the job I'm at even if I'm unhappy there.", and another states "I do not like feeling solely responsible for all of our financial needs." (Ford, 2017). Similiar statements were made by German women and men interviewed for a coverage on the persistence of the male breadwinner model and its financial consequences, published by the German magazine *Der Spiegel*. For example, one man says: "During the period in which I spent all my time taking care of our daughter, my wife told me that being solely responsible for our family's financial situation weighs heavy on her shoulders." (Dahlkamp et al., 2017, translated by L.G.)

or home production (labelled neutrally as "paid" and "unpaid" task during the experiment). The production process takes five minutes. Upon its completion, all income generated by the two players is pooled and each player receives half of the total household income. Note that I impose this sharing rule exogenously, implicitly setting the weight of the female utility in the couple's welfare function $\alpha = 1$, to focus on the labour supply decision and avoid creating a hold-up problem (Hart, 1995).[29] Pooling also ensures that, for any possible labour division, supplying maximum effort is an income-maximising strategy.

As in the model discussed in Section 2, only market work generates income for the household, while working in the home produces a household public good, which generates additional utility. Consequently, if both players engage in market work, each of them generates income for the household according to their individual productivity. Individual market "wages" are given by $w_i = xn_i \forall i \in \{m, f\}$, where x is a fixed piece rate for each item produced, and n_i is the number of items player i produces within the production time. If one partner, say j, engages in household production rather than working in the market, j does not generate income, but i's wage rate is doubled, i.e., $w_i = 2xn_i$. This ensures that the key features of the two activities—market work is paid, home production is unpaid—are preserved, while at the same time, household production adds to the overall consumption level of the household. For i's wage rate to be doubled, j is required to supply effort; the minimum number of items required is derived from j's performance in a shorter, individual production phase in a previous stage of the experiment, see the following subsection for details.[30] This feature is important because it guarantees that productive efficiency is maximised by specialising according to *absolute* rather than *comparative* advantage in market production.

[29]Leaving the decision how to split income to the player who earns it, may lead to strategic reservations for the player who would optimally specialise in home production. The hold-up problem was studied experimentally by Güth et al. (2003, 2004); Oosterbeek et al. (2003).

[30]Note that, conditional on performing the unpaid task, income pooling makes its completion a dominant strategy. This is corroborated by the data; not a single subject did not produce the minimum number of items required to double their partner's wage rate.

Players can choose one of three options, labeled neutrally as Options 1, 2, and 3. Option 1 is a *dual earner labour division*, in which both players work in the market, each contributing w_i to total household income. Option 2 is a *female earner labour division*, in which f works in the market and generates a total household income of $2w_f$, while m produces the household good. Finally, Option 3 is the *male earner labour division* and symmetric to Option 2. Table 4.2 summarises the available options and the resulting individual payoffs. Clearly, the three are pay-off equivalent if m and f are equally productive in market work. Whenever productivity differences occur, however, choosing the partner who holds a productivity advantage in market work as single earner maximises pay-off.

Table 4.2: Payoffs for different choices of labour division

Labour division	Paid work	Unpaid work	Payoffs $\pi_i \forall i \in \{m, f\}$
Option 1: Dual earner	m, f	—	$\pi_i = \frac{w_m + w_f}{2}$
Option 2: Female earner	f	m	$\pi_i = w_f$
Option 3: Male earner	m	f	$\pi_i = w_m$

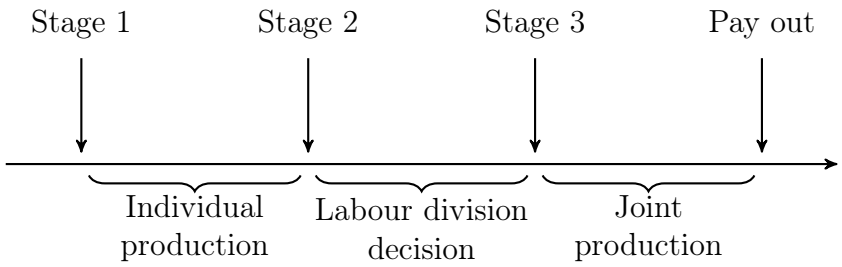
4.3.2 The tasks

The production process requires subjects to engage in real effort tasks. Wage income is produced by performing a task supplied by Benndorf et al. (2014), which is a modified version of the Encryption task designed by Erkal et al. (2011). The household good is produced pby erforming the Slider task designed by Gill and Prowse (2012). The tasks were chosen to minimize systematic gender differences in productivity favouring men, and beliefs about such differences. Prior studies using these tasks have documented either no gender differences in performance, or a slight advantage for female players, see Gill and Prowse (2014) and Gerhards

and Siemer (2016) for the Slider task, and Dato and Nieken (2014) and Benndorf et al. (2014) for the Encryption task.

Subjects' baseline productivity in both tasks was measured in stage 1, before they learned about the specialisation game. No interaction with other participants took place in this stage; subjects simply played each task individually over three one-minute rounds (see Figure 4-1). The first trial round was unpaid; the second paid a piece-rate per completed item, which was doubled in the third round. Round 3 wage rates in stage 1 were $w_i = 2xn_i$, which corresponds to the piece rate subjects later received in stage 3 when working as a single earner in the specialisation option.

Figure 4-1: Progression of the experiment



Stage 1 served several important purposes. First, subjects gained experience with the two tasks, which became important at the subsequent stage when they were asked to choose a labour division. Second, they received feedback about their absolute productivity (number of items completed per minute), and discovered that the scope for learning is rather limited in both tasks.[31] Third, participants were asked to predict their own relative rank among all participants in their session as well as their partner's rank, and were also asked to predict the number of female participants among the top half of performers, which could then be used to verify that participants did not expect gender differences in performance. Finally, stage 1 performance on the Slider task (round 3) determined the target number of Sliders: in order to double their partner's pay rate, a subject working in home production

[31]Stage 1 productivity increases from round 2 to round 3 were small, less than one item on average, see Appendix Figure 4-9.

during the joint production process in stage 3 would have to complete at least as many Sliders as he or she previously had during the individual production stage 1 (round 3). This design feature guaranteed that homemakers would have to supply a certain amount of effort, while at the same time eliminating the risk of failing despite their best efforts. Eliminating risk of failure is important to not increase the expected value of the dual earner option relative to the single earner options. Furthermore, previous research suggests that women are more risk-averse than men (Charness and Gneezy, 2012; but see also the critical assessments by Nelson, 2015; and Filippin and Crosetto, 2016). Therefore, eliminating the risks associated with the single earner options is important to mitigate gender differences in labour division choices driven by risk aversion but unrelated to identity concerns.

4.3.3 Labour division choices and their execution

At the beginning of stage 2, subjects received instructions with the rules of the Specialisation game (see Appendix 4.6 for a translated version). Before making any decisions, participants completed a paper-and-pencil test to demonstrate their comprehension of the game and the pay-off consequences of the three options.[32] Upon completion of the test, subjects were matched with a partner (either their real partner or a randomly assigned stranger of the same sex as the partner), and were asked to choose their preferred labour division. Within a pairing, each subject made three choices: The first two were made completely autonomously from the partner, once without information on his or her partner's productivity, and once with the partner's stage 1 productivity (round 3) displayed on the screen, alongside subjects' own productivity in that same round. The third decision was made jointly by the partners in a chat room.[33]

[32]Lab assistants were instructed to collect and check the tests, and to resolve comprehension issues with those participants who were having trouble solving the test on their own. A translated version of the test can be found in Appendix 4.6.

[33]This chapter focusses on analysing individual decisions.

Subjects were informed that one of the choices made by themselves or their partner (i.e., a total of five per pairing: two individual choices per partner plus their joint choice) would be drawn at random, and then executed and paid accordingly. Before learning the result of the draw, subjects were rematched and repeated the previous steps with a different partner. Finally, subjects were paired with five different computer partners of varying productivity.[34] Information on computer productivity was available immediately. The instructions stated that computer partners, when performing the paid Encryption task, would solve precisely the number of items per minute as indicated by their productivity. As performers of the unpaid Slider task, they were guaranteed to complete the number of Sliders required to double the piece rate of the subject performing the paid task. Again, one of the five choices regarding a computer partner would be randomly selected for execution. Consequently, stage 3 of the experiment took about 15 minutes, during which subjects successively underwent the joint production process with the three different partners. Each time, they were informed which labour division option was selected, and thus which task they were going to perform, just before production began.

4.3.4 Recruitment and sample

The experiment was programmed in zTree (Fischbacher, 2007) and conducted in the WiSo Experimental Laboratory at the University of Hamburg, Germany, in 2016 and 2017. A total of 246 subjects participated in 12 sessions. The median session comprised 20 participants, with 12 in the smallest and 28 in the largest session. Subjects were recruited via hroot (Bock et al., 2014) from a regular student subject pool. Sessions lasted about 2.5 hours and average payouts were €26 (about $32.83 at the time of writing this chapter). The invitation email required subjects

[34]The order of playing real partner/stranger was varied across sessions, computer partners were always the third match. Since participants did not learn about the labour division decision that was selected with a specific partner before all choices in all pairings had been made, the order of play is unlikely to affect choices.

to register and bring their partner to the experimental session. Being married or cohabitation was not a prerequisite, nor was it required that the partner be enrolled as a student.

Upon arrival in the lab, participants were reminded that given the purpose of the study, only real couples could participate (i.e., not study partners or housemates). They were asked to leave the session if they had not come with their real partner, but no one did. To alleviate any potential concerns about "fake couples" contaminating the analyses presented in the following sections, I use a battery of questions from the post-experimental questionnaire to cross-check partners' answers for consistency. These items include questions about the partner's date of birth and where they first met. Using a within-couple consistency score, I find that answers are 82.8% consistent in the median couple. Only 10 couples achieve less than 50% consistency, and excluding them from the analyses presented in this chapter does not alter the results. The complete set of questions and details regarding the procedure can be found in Appendix 4.7.

4.3.5 Summary statistics

Table 4.3 presents summary statistics by gender.[35] The variables are obtained from the post-experimental questionnaire and grouped into four categories: Participant and union characteristics, current enrollment into an educational institution, current activity in the labour market, and experience with economic experiments or economic coursework. As can be seen from the test statistics shown in the last column, most differences between men and women are statistically insignificant,

[35]The data analyses and the presentation of results in this chapter were prepared using the *R programming environment for statistical computing* (R Core Team, 2018), version 3.5.0., and several add-on packages: *broom* (Robinson, 2018), *car* (Fox and Weisberg, 2011), *lme4* (Bates et al., 2015), *lsmeans* (Lenth, 2016), *lmtest* (Zeileis and Hothorn, 2002), *multcomp* (Hothorn et al., 2008), *multiwaycov* (Graham et al., 2016), *plm* (Millo, 2017), *stargazer* (Hlavac, 2018), *sandwich* (Zeileis, 2006), *tidyverse* (Wickham, 2017), *xlsx* (Dragulescu, 2014), and *xtable* (Dahl, 2016). For calculating cluster-robust confidence intervals, I use a function written by Gubler (2014).

Table 4.3: Summary statistics by gender

	Female (F)		Male (M)		Difference (F-M)	
Variable	N	Mean	N	Mean	Mean	p-value
Age*	123	24.89 (4.65)	123	26.24 (4.66)	-1.35	0.023
Union duration*	122	3.01 (3.18)	122	3.00 (3.55)	0.01	0.976
Cohabitating duration*	37	2.89 (3.81)	38	2.98 (4.31)	-0.09	0.925
Married	123	0.03 (.)	123	0.03 (.)	0.00	1.000
Satisfaction with union*	123	8.78 (1.56)	123	8.69 (1.73)	0.09	0.67
Not in education	123	0.12 (.)	123	0.15 (.)	-0.03	0.579
In school	108	0.03 (.)	104	0.04 (.)	-0.01	0.96
Undergraduate student	105	0.59 (.)	100	0.65 (.)	-0.06	0.463
Postgraduate student	105	0.28 (.)	100	0.23 (.)	0.05	0.549
Other student	105	0.13 (.)	100	0.12 (.)	0.01	0.939
Not active in the labour market	123	0.30 (.)	123	0.33 (.)	-0.02	0.783
Full time employed	123	0.14 (.)	123	0.20 (.)	-0.07	0.236
Part-time employed	123	0.20 (.)	123	0.20 (.)	0.01	1.000
Irregularly employed	123	0.36 (.)	123	0.28 (.)	0.08	0.218
Unemployed	123	0.04 (.)	123	0.04 (.)	0.00	1.000
Ever studied economics	123	0.36 (.)	123	0.46 (.)	-0.11	0.12
Ever participated in experiment	123	0.62 (.)	123	0.56 (.)	0.06	0.437
Ever played Slider task	123	0.21 (.)	123	0.16 (.)	0.05	0.414
Ever played Encryption task	123	0.09 (.)	123	0.04 (.)	0.05	0.196

Note: Variables are contious when asterisked and binary otherwise. p-values are obtained from a test of equal proportions for binary, and from a t-test for continuous variables. Standard deviations are reported in parantheses for the means of continuous variables. Age and union duration are measured in years, satisfaction with the union is self-reported on a 11-point scale, where 0 indicates not satisfied at all and 10 very satisfied.

with the exception of age: men are about 1.35 years older on average, and the difference is significant at the 5% level. Note that this tendency for men to be older than their female partners reflects a general pattern in the population of heterosexual couples in Germany (and many other countries of the world).[36] The youngest (oldest) participant was 17 (48) years old and the median was 25.

On average, couples had been together for 3 years at the time of the experiment. The longest duration was almost 27 years, the shortest was less than one

[36]The German statistical office reports that in 2014, men were older than their partner in 73% of couples, and women in only 17% (Krack-Roberg et al., 2016).

month, and the median relationship length was 1.8 years.[37] Among those who lived together (nearly one-third), the average cohabitation time was close to 3 years. Only 3% of the participants were married to each other and no couples had children (3 subjects reported having children with a former partner). The average reported satisfaction with the relationship was very high—nearly 9 on an 11-point Likert scale—where 10 indicates the highest possible level of satisfaction.[38]

With regard to subjects' current engagement in education, the majority were enrolled in some type of educational program; less than 15% were not. A very small fraction was currently attending high school or receiving vocational training, and the rest were studying at university. Students predominantly consisted of undergraduates (bachelor's level), with around 25% postgraduate students (master's or PhD level), and 11% in some other category.[39] Nearly one-third of subjects were not active in the labour market at all. The rest were predominantly part-time or irregularly employed (about 50%), with 17% reporting full-time employment, and a small minority (4%) being unemployed.

subjects' experience with economics and economic experiments, as this may affect labour division choices. Around 41% of subjects reported having taken an economics class at some point. The difference between men and women amounts to nearly 11 percentage points, but just fails to meet statistical significance at the 10% level. The share of those who reported having ever participated in an economic experiment is nearly 60%. Since partners' "work experience" in the specific tasks used in this experiment could affect which partner is chosen as breadwinner, it is important to note that less than one out of five subjects had previously encountered

[37] A boxplot showing the distribution of couples over relationship length can be found in Appendix-Figure 4-8. All results presented in this chapter are robust to the exclusion of couples who were togther for less than three months, which is about 10% of the sample.

[38] A positive selection of couples regarding (self-reported) satisfaction with their relationship has also been noted in other family experimental studies that have been conducted in developed countries (e.g., Beblo and Beninger, 2017).

[39] The German university system occasionally still features diploma degrees, and state examination degrees that have not been integrated into the bachelor's/master's system—mostly law and medical degrees.

the Slider task, and less than one out of ten the Encryption task. If anything, women were slightly more experienced in both tasks, but neither difference is statistically significant.

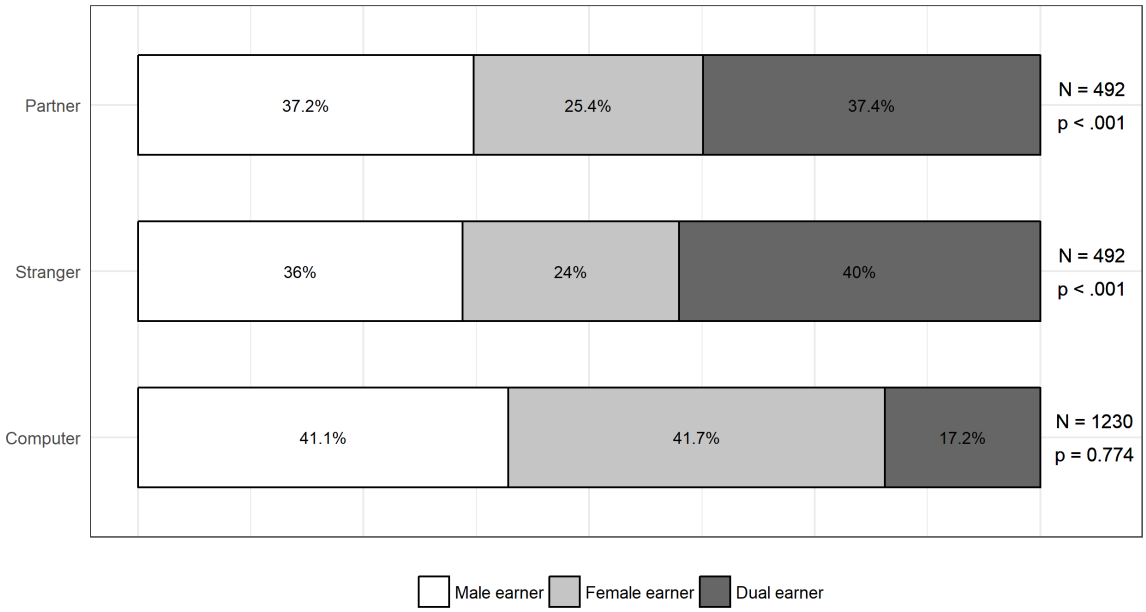
4.4 Empirical results

4.4.1 Labour division choices

We begin by examining the labour division choices broken down by treatment, pooling all autonomously made individual decisions. This includes, per subject, two choices per human partner (real partner and stranger) and five choices for different computer partners of varying productivity, i.e., a total number of $9 \times 246 = 2,214$. To make the different partner treatments comparable, all choices are coded from the subject's perspective, regardless of the pairing. Thus, when subjects are paired with a computer, the classification of a given labour division choice as male or female earner depends on the subject's gender: For a male subject, choosing the computer (himself) as single earner is classified as a female (male) earner, as would be the case when the male subject is paired with a human partner. Women's choices are coded reversely.

Figure 4-2 shows the relative frequency of the three labour division options in each partner treatment. In human pairings, the dual earner labour division is chosen most frequently, followed by the male earner labour division. Female earners are chosen in roughly one in four cases. A test of equal proportions of male and female earners is rejected at the 1% level in both treatments with human partners. When paired with a computer, however, the share of dual earner choices decreases markedly, and male and female earners are chosen at about the same rate (proportions do not differ statistically). Given the coding of the computer choices, the increase in the share of female earners could be driven by women being more willing to choose themselves as earners when playing with a computer, men being

Figure 4-2: Labour division choices across treatments



Note: Proportions calculated from the raw data, p-values obtained from a test of equal proportions comparing female and male breadwinner choices.

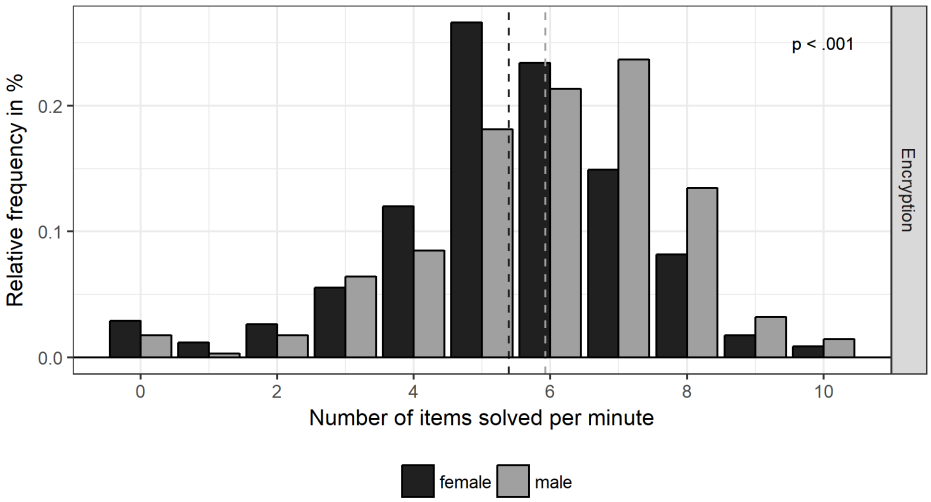
more willing to choose their partner as earners, or both. A disaggregated analysis by gender is presented in the next subsection.

The pattern shown in Figure 4-2 may indicate that identity concerns influence subjects' labour division decisions, because gender differences in the probability of being nominated as breadwinner are only present in human partners, but entirely absent in pairings with computer partners, for which no social prescriptions exist. However, this simple comparison does not account for intra-couple productivity differences, which may be driving these results. Despite the fact that both tasks have repeatedly shown to be gender-neutral in previous research, or to even favour women slightly, the women in this sample were, on average, less productive than men when baseline productivity measures were collected in stage 1. As Figure 4-3 shows, stage 1 productivity in the Encryption task differed significantly between male and female subjects at the 1% level. On average, men encrypted 5.9 items per minute, 0.5 more than women. In light of these differences, it is possible that the overrepresentation of male earner choices merely reflects income-maximising behaviour and is unrelated to identity concerns. The next subsection will therefore present a more sophisticated approach to disentangle the two.

4.4.2 The role of income maximisation and identity motives

As discussed earlier, income maximisation and identity goals are aligned when the male partner holds a productivity advantage, but in conflict when it is held by the female partner. Consequently, if identity concerns exist, individuals in the latter type of couple should be relatively less likely to make an income-maximising choice. I test this idea by studying the share of subjects making an income-maximising choice, comparing the marginal effect of the female vs. the male partner holding a productivity advantage. For the dependent variable, I construct an indicator, inc^{max}, that is equal to 1 if a labour division choice is ex-ante income-maximising, given partners' stage 1 productivities in the Encryption task. Note that, in the absence of productivity differences, all three labour divisions are ex-ante pay-off

Figure 4-3: Male and female stage 1 productivity



Note: Performance data pooled from 3 rounds in stage 1. Dashed lines show means, p-values are obtained from a Kolmogorov-Smirnov test for the equality of distributions.

equivalent; thus, the indicator always equals 1 in these pairings. If one partner was more productive in stage 1, only choosing that partner as wage earner and the other as homemaker is coded as income-maximising. According to this classification, around 37% of choices are not income-maximising when paired with a human, and 16% with a computer partner. They predominantly comprise the dual earner option, which accounts for three-quarters of the non-income-maximising choices concerning a human, and 60% involving a computer partner. Thus, subjects rarely choose a partner with a productivity disadvantage in the market task as the single earner.

Exploiting the panel structure of the data allows accounting for multi-level clustering at the individual and session levels. I estimate the probability that

individual i in session s makes an income-maximising choice in round r, using the following linear probability model:

$$\begin{aligned} inc_{isr}^{max} =& \beta_0 + \beta_1 P_{ir}^{male} + \beta_2 P_{ir}^{male} \times Stranger_{ir} + \beta_3 P_{ir}^{male} \times Computer_{ir} + \\ & \beta_4 P_{ir}^{female} + \beta_5 P_{ir}^{female} \times Stranger_{ir} + \beta_6 P_{ir}^{female} \times Computer_{ir} + \\ & X_{ir} + u_i + v_s + \epsilon_{isr} \end{aligned} \tag{4.9}$$

In this multi-level model, u_i is the subject-specific and v_s the session-specific random effect; ϵ_{isr} is the usual error term, with the standard assumptions regarding zero mean and variance (Moffatt, 2015).[40] The vector X_{ir} holds a set of individual-specific controls that change over rounds, i.e., the absolute difference between partners' stage 1 productivity in the paid task and its square, as well as partners' stage 1 productivity in the unpaid task, which are included in some specifications. Since all choices are considered ex-ante income-maximising when partners are equally productive, the reference category includes all pairings in which this is the case (partner, stranger, or computer), and β_0 should be estimated as 1. P^{male} is an indicator that is equal to 1 when the male partner holds a productivity advantage. Since its interactions with the indicators *Stranger* and *Computer* are also included, the coefficient β_1 measures how much the probability of choosing an income-maximising labour division declines in *real couples* when the male partner holds a productivity advantage, relative to pairings without productivity differences. β_2 and β_3 measure how the effect changes, respectively, in pairings with strangers and computers. The coefficients on the indicator P^{female}, and its interactions with *Stranger* and *Computer*, measure the corresponding effects of a productivity advantage held by the female partner. To test the identity hypothesis that a productivity advantage affects income maximisation differen-

[40] Note that I have estimated the model additionally including a couple-specific random effect to allow for a further level of clustering. The results are virtually indistinguishable, and a likelihood ratio test fails to reject the null hypothesis that the three-level model presented here is nested within the four-level model.

tially when held by the male partner vs. the female partner, I conduct linear hypothesis tests of the null that the two effects are equal.[41]

Table 4.4 supplies the first set of regression results, using only individual choices under full information about stage 1 productivity differences in the paid task. The first column presents coefficients from a restricted model, in which the indicator P is equal to 1 if any partner holds a productivity advantage. Columns (2)-(4) show coefficients from the fully specified model for the full sample, as well as for men and women separately. By construction, the intercept, which measures the probability of choosing an ex-ante income-maximising labour division in the absence of productivity differences, is close to 1 across all models. The first important insight we gain from the restricted model (column 1) is that the probability reduces by almost 36% in real couples with productivity differences, as the negative and highly significant coefficient on P shows. We also learn that decisions are significantly more likely to be income-maximising in pairs of strangers and in pairings with computers. Both differences, around 10 and 16 percentage points, respectively, are statistically significant at the 1% level.

As noted above, most of the reduction in the probability of choosing an ex-ante income-maximising labour division in the presence of productivity differences stems from dual earner option choices. Several reasons might explain this option's popularity. First, it is possible that subjects perceive the stage 1 productivity measure as an inaccurate predictor for their partner's prospective stage 3 contribution to joint household income. For computer partners, however, the measure is a perfect predictor for their prospective stage 3 productivity. Therefore, a higher rate of *inefficient*[42] dual earner choices in pairings with a human partner could reflect a

[41] For real partners, this requires testing that $\beta_1 - \beta_4 = 0$. For stranger and computer partners, the difference in the linear combinations of the productivity indicator and its respective interaction are tested, i.e., $\beta_1 + \beta_2 - (\beta_4 + \beta_5) = 0$ for strangers, and $\beta_1 + \beta_3 - (\beta_4 + \beta_6) = 0$ for computers.

[42] For convenience, I use the term "inefficient" in the narrow sense of productive efficiency, as the more precise description of a choice as "not income-maximising" at times reduces language efficiency beyond a tolerable degree. I do, however, acknowledge, that income-maximisation may

Table 4.4: Proportion of income maximising choices, *with* information on partner productivity

	Choice: Income maximising			
	(1)	(2)	(3)	(4)
Intercept	0.999***	0.999***	0.996***	1.006***
	(0.020)	(0.020)	(0.029)	(0.030)
PA	−0.355***			
	(0.025)			
$PA \times Stranger$	0.101***			
	(0.025)			
$PA \times Computer$	0.156***			
	(0.021)			
PA^m		−0.349***	−0.254***	−0.370***
		(0.028)	(0.051)	(0.050)
$PA^m \times Stranger$		0.141***	0.023	0.160***
		(0.032)	(0.063)	(0.062)
$PA^m \times Computer$		0.129***	0.063	0.124***
		(0.026)	(0.048)	(0.046)
PA^f		−0.369***	−0.240***	−0.432***
		(0.036)	(0.067)	(0.066)
$PA^f \times Stranger$		0.050	−0.115	0.207**
		(0.042)	(0.082)	(0.081)
$PA^f \times Computer$		0.195***	0.040	0.303***
		(0.035)	(0.064)	(0.065)
Productivity controls	N	N	N	N
Information	Y	Y	Y	Y
Sample	All	All	Men	Women
Observations	2,214	2,214	861	861

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01. Sample includes only decisions taken with information on partners' stage 1 productivity. Estimations including controls for the absolute productivity difference between partners and its square can be found in Appendix 4.8.

hedging strategy. Notably, though, the increase in the rate of income-maximising choices when paired with computers compared to the increase observed in pairings of strangers does not differ statistically. Therefore, a second reason for why ex-ante income maximisation, given productivity differences, is lowest in real couples might be an elevated preference for the dual earner option when playing with one's real partner. This might arise if it is perceived as more "fair" or "fun" to play dual earner as a couple, e.g., because both partners contribute to wage production, or, if there is altruism between partners and the unpaid Slider task is perceived as less enjoyable than the Encryption task. When playing with a computer, such concerns are certainly absent, and they may also operate to a lesser extent in pairings of strangers. Finally, a third motive for the dual earner option preference over the income-maximising choice may arise asymmetrically in pairings where the female partner has a productivity advantage, if subjects hold identity concerns and are thus reluctant to choose a female earner.[43] Again, such concerns should not be present when paired with a computer, and may also be attenuated when paired with an anonymous stranger rather than one's real partner.

The impact of identity concerns can be disentangled by examining how much *more* the probability of choosing an income-maximising labour division diminishes when the productivity advantage is held by the female partner versus the male. This is done using the fully specified model presented in columns (2)-(4). Figure 4-4 visualises the separate estimation results for men and women, and additionally reports p-values obtained from the linear combination tests, thus allowing conclusions regarding the statistical significance of the differential response to a male vs.

not be the sole, or even primary goal for subjects, and that their choices may well be efficient given their preferences.

[43]Note that choosing the dual earner option is ex-ante less costly than choosing male earner when the female partner is more productive.

female productivity advantage.[44] The panels in row 1 of Figure 4-4 correspond to columns (3) and (4) in Table 4.4.

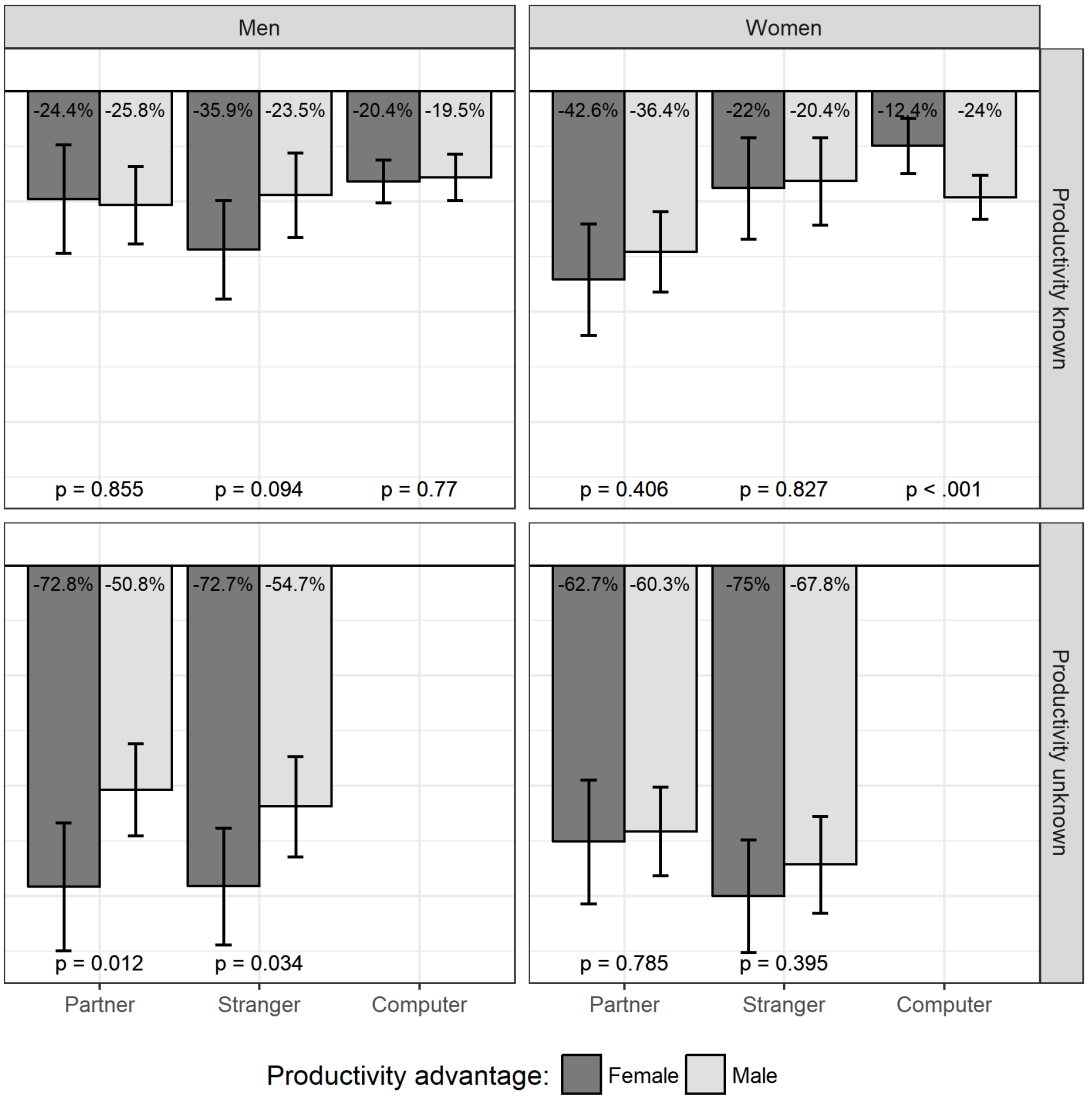
We begin by studying men's decisions. As the top left panel reveals, their willingness to choose an income-maximising labour division is roughly consistent across the different pairings and mostly unaffected by the gender of the most productive partner. The fact that men do not treat a productivity advantage differently when it is held by a woman as compared to when it is held by themselves indicates that they are not motivated by identity concerns.[45] The one situation in which their response is notably asymmetric is when paired with a more productive stranger. In stranger pairings, men are more than 10 percentage points less likely to choose an income-maximising labour division when this requires choosing their female partner as the breadwinner compared to choosing themselves when they are more productive; the difference is significant at the 10% level. While this possibly reflects identity preferences that men seek to hide from their partner, it seems more plausibly related to the fact that uncertainty about partner productivity is highest in stranger pairings, where information about an unfamiliar partner's past productivity may be perceived as a weaker predictor of future productivity. Consequently, there is more room for men's overconfidence to affect their beliefs about future contributions to household income production.

The panel below (bottom left) presents evidence that men's overconfidence indeed has a strong effect on their labour division choices. The results are obtained from regressions similar to those presented in Table 4.4, but only include decisions for which information on partners' stage 1 productivity was not yet available. Hence, computer pairings are excluded, as computer partners' productivity was al-

[44] All results presented here are obtained from the baseline model without further controls. Appendix Tables 4.8 and 4.9 show that the results are robust to the inclusion of the productivity controls.

[45] Note that, while I argue most of the time that the identiy prescription is "A man should earn more than his wife," it may of course entail other behavioural prescriptions, such as a man should behave like a gentleman, i.e., take the work upon himself in order to save his spouse the trouble. However, such motives should also lead men to choose qualified women at lower rates as breadwinners, and thus they, too, are not supported by the data.

Figure 4-4: Marginal effects of market productivity advantage on the probability to choose an income-maximising labour division



Note: Marginal effects, p-values, and 95% confidence intervals (shown by whiskers) are calculated from the linear models (3) and (4) in Table 4.4 for the top row (productivity known) and Appendix-Table 4.7 for the bottom row (productivity unknown).

ways known. Not surprisingly, labour division choices are overall much less likely to be income-maximising compared to rounds where information is available. This means that, to some extent, these choices include "honest mistakes" and do not reflect true preferences for a particular labour division. Interestingly, men make mistakes at a relatively higher rate when the female partner is more productive as compared to when they hold the productivity advantage themselves. In both real partner and stranger pairings, men are about 20 percentage points less likely to choose the income-maximising labour division when their female partner is more productive than when they are. The differences are significant below the 5% level. Both findings together—men's behaviour when partner productivity indicators are unavailable, or available but less conclusive, as in the case of strangers—indicate that men are less likely to choose a qualified woman as a breadwinner relative to a qualified male, when information on relative productivity differences is noisy. This indicates that male overconfidence may not only explain why men are overrepresented in top-level jobs, as argued by Niederle and Vesterlund (2007), but also why they are more likely to assume the role of the main earner in their families.

Turning to women's choices in the second column of Figure 4-4, the pairing appears to matter for income maximisation on two dimensions when information on partner productivity is available (top right panel). First, choices are more likely to be income-maximising when women are paired with strangers or computers as opposed to their real partners (all increases are significant below the 5% level, as seen in Table 4.4). Second, there is heterogeneity in women's responses to the gender of the partner holding the productivity advantage across pairings. While we see no differential treatment of a male productivity advantage vs. a female when women are paired with human partners, a distinctly asymmetric effect emerges when women are paired with computers. Here, women's choices are significantly more likely to be income-maximising when women themselves hold a productivity advantage rather than the computer partner. Comparing the effect of the female productivity advantage across pairings also reveals that, when qualified for the

job, women choose themselves as breadwinners at a much higher rate when paired with computers relative to being paired with real men. It appears that women shy away from assuming the breadwinner role when paired with their partners, but for reasons that may be unrelated to identity. One such potential reason is the perceived pressure of performing as a single versus dual earner, which will be explored in more detail in the next subsection.

We conclude this subsection by studying the bottom right panel, which displays decisions made absent of information on the partner's stage 1 productivity. While women, like men, are also generally more likely to make mistakes than when information is present, the rate of mistakes does not vary with the gender of the productivity advantage holder. Therefore, unlike men's, women's mistakes do not appear to be related to their confidence.

4.4.3 Women shying away from the provider role?

To investigate the hypothesis that women anticipate stronger feelings of pressure as sole income providers, I closely follow a procedure used by Dohmen and Falk (2011) in their seminal paper on the role of productivity, preferences, and gender for sorting decisions into different payment schemes. The authors had subjects perform a task under a piece rate in an introductory stage, then asked them to choose a payment scheme under which they would like to perform the task again, comparable to the setup in this chapter (individual performance in stage 1, sorting into labour divisions in stage 2). Employing a similar strategy, I asked subjects to rate the tasks on several dimensions upon completion of stage 1, but before confronting them with the labour division decision. I adapt the set of questions used by Dohmen and Falk (2011) slightly, so as to better match the labour division decision investigated in this chapter.

Specifically, subjects are asked to rate the tasks in terms of how (i) entertained, (ii) stressed, (iii) bored they were during the stage 1 performance, and (iv) how

much effort they exerted.[46] After having made their labour division choices in stage 2, but before learning the outcome of the random draw and actually performing the selected labour division in stage 3, subjects are asked to imagine that a labour division will be drawn that requires them to perform the paid Encryption task. Half of the sample is asked to imagine this because the dual earner option is selected (i.e., both the subject and their partner perform the paid task). The other half is asked to imagine that they perform as single earners, while their partner performs the unpaid task.[47]

To study potential differences in the anxiety levels men and women anticipate when performing the same task, i.e., the Encryption task, as single vs. dual earners, I estimate the following equation:

$$Y_i = \beta_0 + \beta_1 Single\_earner_i + \beta_2 fem_i + \beta_3 fem_i \times Single\_earner_i + \boldsymbol{X_i} + \epsilon_{ic} \tag{4.10}$$

Y_i is one of the four outcome variables: the level of entertainment (fun), stress, or boredom i expects to experience when performing the task, as well as how much effort he or she expects to exert (measured on a 1-7 Likert scale, where 1 means "not at all" and 7 "very much"). $Single\_earner_i$ is an indicator that equals 1 for individuals who were asked to imagine performing the task as single earners, and zero for those who were asked to imagine performing as dual earners. fem_i is an indicator for female subjects. The vector $\boldsymbol{X_i}$ includes control variables for subjects' stage 1 performance in the task, as well as their partner's. Most importantly, it

[46] Dohmen and Falk (2011) ask subjects to rate the task on three dimensions: How much effort they exerted, how stressed they felt, and how exhausted they felt.

[47] Subjects are asked for these prospective ratings twice: Once after they made all choices with their real partner, and with the randomly matched stranger. For the estimation results presented below, I pool all ratings, but the results are qualitatively unaffected when excluding ratings that refer to a labour division with a stranger.

includes subjects' *retrospective* stage 1 assessment of the task with respect to the same dimension, i.e., how much fun, stress and boredom they experienced or how much effort they exerted, respectively, during their performance of the task in stage 1. Thus, the coefficient on β_1 measures how performing the task as single earner changes men's anticipated response. β_2 captures any difference in the response of women who anticipate performing as dual earners relative to men in the same role. The coefficient on the interaction term, β_3 captures the difference in differences of the single vs. dual earner roles across genders. Consequently, the linear combination of β_2 and β_3 measures the difference in anticipation between male and female single earners.

The results are summarised in Table 4.5. Column (1) refers to the level of stress subjects anticipate experiencing in stage 3 if the random draw selects them as the paid task performers. As hypothesised, women do not anticipate experiencing stress levels different from men as dual earners, but their anticipated stress levels increase significantly when they imagine themselves performing as single earners. Because the linear combination of β_2 and β_3 is significant at the 5% level, we can conclude that women are more stressed by the prospect of performing as single earners than men. A symmetric pattern unfolds in column (2) with respect to boredom. Here, higher values indicate that a subject anticipates greater levels of boredom during stage 3. Thus, women expect higher levels of boredom in the role of dual earner than men in the same role, as shown by the negative coefficient on fem, but the effect reverses when they imagine themselves as single earners. Again, the linear combination of β_2 and β_3 is significant, here at the 10% level, which supports the conclusion that women expect lower levels of boredom when performing as single earners compared to men in that role. Taken together, the two results on stress and boredom point to elevated levels of anxiety associated with the breadwinner role for women but not for men, and may explain why women are so much more likely to choose themselves as breadwinners when paired with a computer than with a human partner.

Table 4.5: Prospective feelings about the paid task

	Dependent variable:			
	Stress	Boredom	Effort	Fun
	(1)	(2)	(3)	(4)
$Single\_earner$	-0.337	-0.145	-0.403^{***}	-0.079
	(0.243)	(0.206)	(0.151)	(0.233)
fem	-0.112	0.376^{**}	-0.175^{**}	-0.103
	(0.222)	(0.189)	(0.085)	(0.185)
$Single\_earner \times fem$	0.806^{**}	-0.643^{***}	0.316^{*}	-0.091
	(0.340)	(0.248)	(0.161)	(0.315)
$Intercept$	3.348^{***}	2.165^{***}	5.179^{***}	2.386^{***}
	(0.512)	(0.430)	(0.539)	(0.518)
Controls	Y	Y	Y	Y
Sample	All	All	All	All
Observations	492	492	492	492
R^2	0.315	0.289	0.181	0.345

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are clustered at the couple level. All estimations including controls for own and partner productivity, as well as stage 1 assessments of the task. Results are very similiar when excluding assessments that refer to labour division with strangers.

The results for anticipated effort supply presented in column (3) appear to corroborate this conclusion. While women, in the role of dual earners, anticipate supplying less effort than men in the same role, their anticipated effort level is higher when they imagine performing as single earners. Although the linear combination of β_2 and β_3 fails to meet statistical significance at conventional levels in this case, the pattern is qualitatively similar to anticipated stress and boredom. A second important observation is men's strong reaction to the single earner status, for which they anticipate expending *less* effort as compared to the role of a dual earner. In other words, men expect to exert more effort when their partner is also performing the paid task and thus co-contributing to household income. The next subsection will investigate whether this may reflect an identity-related fear of being out-earned by the partner. Finally, column (4) shows that the level of entertainment, or fun, subjects anticipate experiencing during the stage 3 performance of the paid task does not vary across gender, and is unaffected by the single earner status.

4.4.4 Identity and effort supply

In this subsection I study subjects' effort supply in the production stage 3, to investigate possibly heterogeneous responses to stage 1 productivity differences across pairings, which may be related to identity concerns. If subjects are motivated by the social prescription that "a man should earn more than his wife," we may observe male subjects increase and/or female subjects decrease their effort supply in stage 3 if their stage 1 productivities suggest she might out-earn him, i.e., when she was more productive. To study effort supply, I pool all stage 3

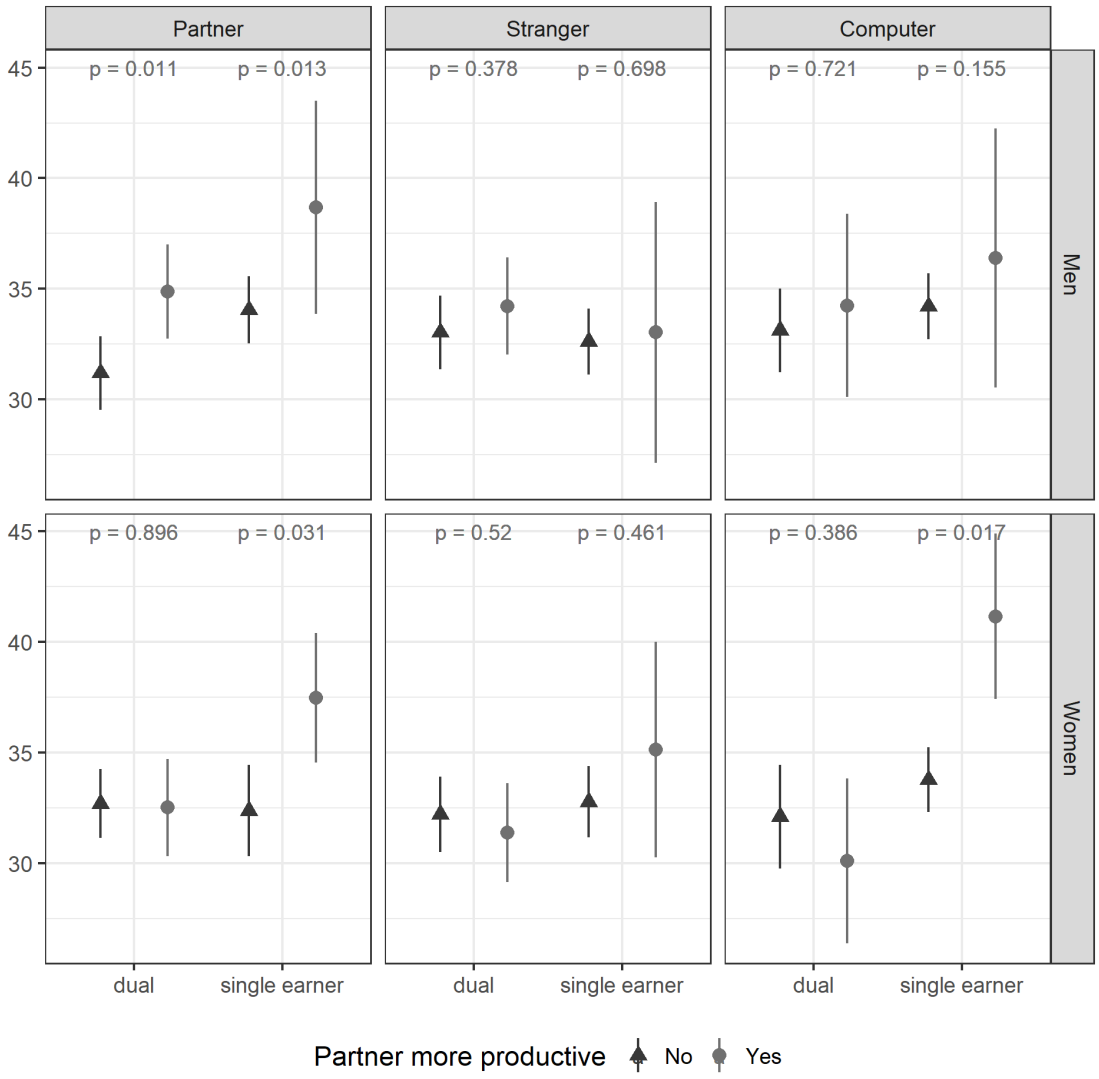
rounds in which individuals performed the paid task either as single or dual earner (462 observations) and estimate the following linear model:

$$Effort_{ir} = \beta_0 + \beta_1 Partner\_more\_productive_{ir} + \beta_2 Single\_earner_{ir} + \beta_3 Partner\_more\_productive_{ir} \times Single\_earner_{ir} + \mathbf{X}_{ir} + \epsilon_i \quad (4.11)$$

where $Effort_{ir}$ denotes the number of items produced by i in a joint production process in round r of stage 3. $Partner\_more\_productive_{ir}$ is an indicator that equals 1 if the partner i is paired with in production round r was more productive in stage 1. Note that, conditional on own and partner stage 1 productivity, which are included in $\mathbf{X}_{ir}$, the realisation of $Partner\_more\_productive_{ir}$ is random. The indicator $Single\_earner_{ir}$ equals 1 if the realised labour division in round r has i perform as a single earner. Thus, dual earners who were equally at least as productive as their partners in stage 1 form the reference group. Recall that each pairing (partner, stranger, computer) has made five potential labour division choices over the course of stage 2 (two autonomous choices by each partner and one joint choice) and only one was randomly selected for execution in stage 3. Thus, conditional on the number of times an individual was chosen as the breadwinner, also included in $\mathbf{X}_{ir}$, performing as a single earner is random. Furthermore, $\mathbf{X}_{ir}$ includes controls for the order of play, and the total number of times a subject performs the paid task in stage 3 that consists of three rounds, one each with the partner, stranger and computer. To ease exposition, the above equation does not lay out the full specification that also includes indicators for female performers, stranger and computer pairings, and their respective interactions, as well as all interactions with the indicators for $Partner\_more\_productive_{ir}$ and $Single\_earner_{ir}$. The full model can be found in Appendix Table 4.12.

The results are visualised in Figure 4-5. Columns 1-3 show the conditional performance averages by pairing; row 1 refers to men and row 2 to women. First, we notice that subjects who were randomly selected as single earners despite their

Figure 4-5: Average effort of paid task performers in stage 3



Note: Conditional estimates obtained from the full model controlling for subject's and their partner's stage 1 productivity, the total number of female and male breadwinner choices, order of play, and the total number of times performing the paid task in stage 3. Whiskers show 95% confidence intervals; p-values are obtained from linear-combination tests of the coefficients from the full model (see column 1 of Appendix-Table 4.12), in which standard errors are clustered at the couple level.

stage 1 productivity succumbing their partner's increase their performance relatively more than single-earners who were at least as productive as their partner in stage 1.[48] For both men and women, this difference is statistically significant when paired with their real partners, but not with strangers. This suggests that they may be motivated to make up for an "erroneously" selected labour division decision, making it ex-post more efficient by exerting high effort. When paired with computers, men and women show a similiar tendency, although the increase in effort supply is significant statistically only in the case of women.[49]

Interestingly, men also increase their effort supply significantly in the dual earner option when their real partner was previously more productive. Specifically, the difference in men's stage 3 performance when their real partner's stage 1 productivity exceeded their own corresponds to about 3 items. The increase is about 10% of the mean performance of men who were not outperformed by their partner in stage 1 and statistically significant at the 1%-level. The fact that their partner's productivity does not affect men's performance when paired with a stranger or a computer points to identity concerns as an explanation. Consistent with the presence of a behavioural prescription that "a man should earn more than his *wife*," men increase their effort supply when the threat of being out-earned by their real partner is high.

4.5 Conclusion

This chapter presented a systematic assessment of the determinants of couples' labour division choices. Motivated by the observation that, despite women's in-

[48] Standard errors are large because this happens rarely.

[49] Recall that subjects made five labour division choices regarding five different computer partners of *varying* productivity. When of of those choices is drawn randomly in stage 3, only the labour division resulting from the draw is revealed. This means that, unless a subject had chosen this particular labour division only for one of the five computers, it is not necessarily clear which computer partner was selected. Therefore, in order to determine whether the computer partner was more productive than subject i, I use the average productivity of all computer partners for which i chose the same labour division.

creasing labour market productivity, only relatively few households choose a female breadwinner, I investigated couples' labour division decision in the lab. Suprisingly, women are significantly less likely chosen as breadwinners than men. Unpacking the mechanisms, gender differences in market productivity are shown to be the main driver of this asymmetry. Moreover, I investigate three other important channels that add to women's underrepresentation among breadwinners.

First, I showed that men's stronger tendency towards overconfidence may be one reason why women are chosen as breadwinners at a relatively lower rate than men, when signals on intra-couple labour market productivity differences are noisy. This mechanism is likely to play an important role in real-world specialisation decisions. This is because, when choosing a breadwinner, comparing expected lifetime productivities—i.e., how careers, and thus wages, will evolve in the future—is more important than simply comparing contemporary wages, which are much less noisy. If, at the same level of productivity, men are more confident than women regarding their prospective wage-earning capacity, this may result in excess male breadwinner labour divisions.

The chapter also illuminates another important channel, which has not received much attention in the literature: women's reluctance to assume sole responsibility for family income. I present two pieces of evidence that support this hypothesis. First, I show that (qualified) women are much more likely to choose themselves as breadwinners when paired with a computer than when paired with a human partner. Second, I study subjects' prospective assessments of the task and find that this behaviour might be related to the anxiety women anticipate to experience in the role of breadwinners when paired with human partners. Conditional on prior performance and previous assessment of the paid task, the provider role is associated with significantly higher anticipated stress levels and lower levels of boredom for women than for men.

The third channel I investigate is a *direct* effect of gender norms, as proposed by Akerlof and Kranton (2000) and Bertrand et al. (2015).[50] The results here are mixed. On the one hand, I do not find evidence that behavioural prescriptions (e.g., "a man should earn more than his wife") impact the labour division decision; while identity theory posits that specialisation choices are less likely to be income-maximising when a productivity advantage is held by a woman rather than a man, this is not borne out by the data. On the other hand, when examining effort supply for randomly selected labour division choices, I find that men exert significantly more effort when the threat of being out-earned by their partners is high, i.e., if she was more productive than him in a previous stage of the experiment in which subjects performed independently. Because there is no similar effect on men's performance when paired with a stranger or computer partner, I conclude that the driving motive for this behaviour is likely identity considerations, i.e., the fear of being out-earned by their "wife".

Understanding the mechanisms that drive the gendered pattern in family labour division choices is important for designing equalising policies. The issue becomes increasingly pressing, as empirical evidence has shown that gender gaps in economic outcomes start to widen with family formation and childbearing, i.e., when the gains to be had from labour division are highest. The results presented in this chapter suggest that policy strategies targetting women's labour market productivity only, may fall short of eliminating the gender gap in labour force participation. Apart from the channels highlighted here, the finding that much of the gender difference in breadwinner status in this experiment is due to gender differences in productivity stands out and suggests that comparative advantage might be indeed an important driver for couples' labour division choices. The fact that women's productivity in the labour market has caught up with men's in many instances, at least prior to the specialisation decision, suggests that gender differences in home

[50]It is likely that both gender differences in self-confidence and feelings of anxiety associated with sole responsibility for family income, are nurtured by social norms regarding the types of behaviours that men and women are repeatedly taught is appropriate for their gender.

productivity might be important. Shedding light on the relation between women's productivity advantage in the home and gender norms empirically may thus be an important task for future research.

4.6 Appendix A: Experimental instructions

4.6.1 Translated instructions

In the following part of the experiment you will make several decisions and work on tasks to earn money. For this purpose, you need to choose a labour division for yourself and your partner. Before you start, the information about which partner you are matched to will be displayed on your screen.

You can choose between the two tasks you have been working on in the previous part of the experiment. In this part, each of you will work for a total of 5 minutes, but you can decide which of the two tasks you and your partner will perform. The rules are as follows:

- The Encryption task is paid with a piece-rate, i.e., you earn a fixed amount of ECU for each successfully encrypted item.
- The Slider task is now unpaid.
- The piece-rate for the Encryption task will be doubled if one partner completes the unpaid Slider task.
- The total income generated by the partners will automatically be split equally between them.

You can choose from among three labour division options. Please note that your partner will also choose one of the three options. After you have made all of your decisions, the computer randomly determines which choice—yours or your partner's—will be applied.

The three options for labour division are the following:

Option 1: Both partners work for pay.

- Both you and your partner perform the paid Encryption task for a total work time of 5 minutes.

- Each person earns 10 ECU for each successfully encrypted item.
- After the work time has elapsed, both incomes are added up and each partner receives half.

Option 2: The female partner works for pay, while the male partner works unpaid.

- The female partner performs the paid Encryption task and the male partner the unpaid Slider task for a total work time of 5 minutes.
- If the Slider task is completed, the female partner earns 20 ECU for each successfully encrypted item.
- If the Slider task is not completed, the female partner earns 10 ECU for each successfully encrypted item.
- The Slider task counts as completed if the male partner adjusts a minimum amount of Sliders correctly. The target number equals the number of Sliders he adjusted in the last round of the previous part within one minute. The target number must be completed within the total work time of 5 minutes.
- After the work time has elapsed, each partner receives half of the income earned by the female partner.

Option 3: The male partner works for pay, while the female partner works unpaid.

- The male partner performs the paid Encryption task and the female partner the unpaid Slider task for a total work time of 5 minutes.
- If the Slider task is completed, the male partner earns 20 ECU for each successfully encrypted item.

- If the Slider task is not completed, the male partner earns 10 ECU for each successfully encrypted item.

- The Slider task counts as completed, if the female partner adjusts a minimum amount of Sliders correctly. The target number equals the number of Sliders she adjusted in the last round of the previous part within one minute. The target number must be completed within the total work time of 5 minutes.

- After the work time has elapsed, each partner receives half of the income earned by the male partner.

4.6.2 Translated comprehension test

The following short test is meant to verify that you understood the instructions for the next part of the experiment. Please consider the following situation and state your answers to the questions below:

- Two people, A and B, encrypt 10 (A) and 15 (B) words within the available time of 5 minutes.

- In the third round of stage 1 (at a piece-rate of 20ECU), A has positioned 5 Sliders correctly in one minute and B has positioned 6.

1. How much money will A and B receive if they choose option 1 in which both perform the paid Encryption task? _____ ECU for A and _____ ECU for B.

2. If they choose the option in which B performs the paid Encryption task and A performs the unpaid Slider task, how many Sliders must A position correctly in order for the task to count as completed? _____ Sliders in _____ minute(s).

3. If they choose the option in which B performs the paid Encryption task and A completes the unpaid Slider task, how much money will they receive? _____ ECU for A and _____ ECU for B.

4. If they choose the option in which A performs the paid Encryption task and B performs the unpaid Slider task, how many Sliders does B have to position correctly in order for the task to count as completed? _____ Sliders in _____ minute(s).

5. If they choose the option in which A performs the paid Encryption task and B completes the unpaid Slider task, how much money will they receive? _____ ECU for A and _____ ECU for B.

4.7 Appendix B: Couple consistency check

To alleviate concerns about "fake couples" contaminating the analyses presented in this study, I use a battery of questions from the post-experimental questionnaire to cross-check partners' answers for consistency. This allows identifying couples in which partners give inconsistent answers frequently. The specific questions used for this exercise are summarised in Table 4.6.

The first set consists of six questions that each subject answered both in reference to themselves and to their partner. To determine whether both partners in a couple answered consistently, a subject's answer in reference to her partner must be compared against her partner's answer in reference to himself, and vice versa. This set includes questions that ask for the birthday of the partner, whether the subject introduced her partner to her parents, how often the subject stays over at her partner's appartment, and each question vice versa.

The second set consists of five questions for which answers can be compared directly across partners; i.e., the benchmark is the partner's answer to the very same question. These questions ask for the date when the relationship began, how or where the partners met, whether they said "I love you" to the other, whether they have personal items in each other's apartments, and finally, if they have ever seriously discussed having children.

To account for the fact that the number of prespecified answers varied across questions affects the likelihood of a consistent answer (the probability of both partners choosing the same answer is higher for questions with fewer options to choose from) I calculate a "consistency score" for each couple in the following manner: partners receive points for each question they answer consistently and the number of points is equivalent to the number of prespecified answers. For example, if a subject chooses the same answer to the question "Have you and your partner told one another that you love him/her?", the couple receives 4 points because the question has 4 prespecified options (I told my partner; My partner told me; We both told each other; and No). For the start date of the union, answers count as

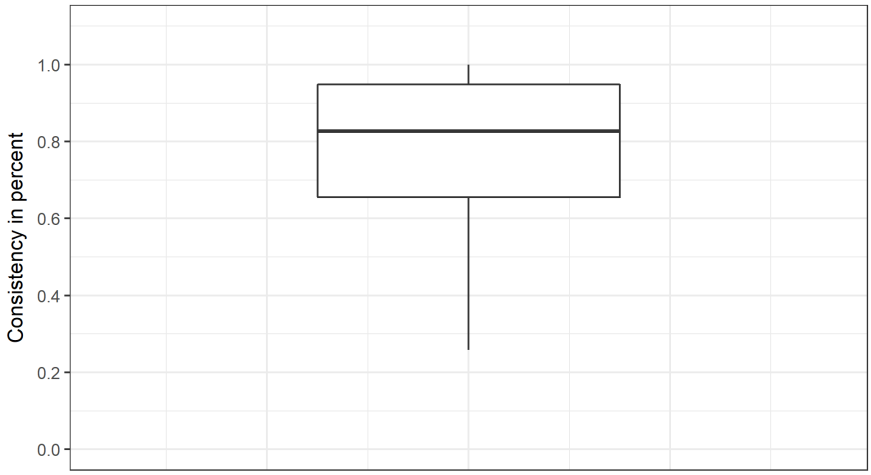
Table 4.6: Questions used to check consistency

Set	Questions	Benchmark	Points
1	Since when (month and year) are you and your partner together?		10
	How did you meet your partner?		10
	Have you and your partner told one another one that you love him/her?		4
	Do you have personal items placed in your partner's appartment or vice versa?		3
	Have you and your partner ever seriously discussed having children?		2
2	Please state your partner's date of birth.	Please state your date of birth.	1
	Please state your date of birth. (+/- 100 days)	Please state your partner's date of birth.	10
	Have you introduced your partner to your parents?	Has your partner introduced you to her parents?	2
	Has your partner introduced you to her parents?	Have you introduced your partner to your parents?	2
	How often does your partner stay over at your appartment?	How often do you stay over at your partner's appartment?	7
	How often do you stay over at your partner's appartment?	How often does your partner stay over at your appartment?	7

consistent and receive 1 point if they do not differ by more than three months. For the open questions on the partner's birthday, couples receive 10 points each only when answered consistently.[51] Note that I do not use the question about the start date for cohabitation to avoid overweighting the many "consistent" missing answers by partners who do not cohabitate. Thus, the maximum consistency score is 58.

The boxplot below shows that the vast majority of couples achieve high consistency rates (obtained by normalising the consistency score). In fact, three quarters of the couples achieve at least 65.5 percent of the maximum score. The median couple achieves 82.8 percent. Only 10 couples (fewer than 9 percent) achieve less than 50 percent of the score. All analyses presented in the chapter are unaffected if these couples are excluded.

Figure 4-6: Couple's normalised consistency score

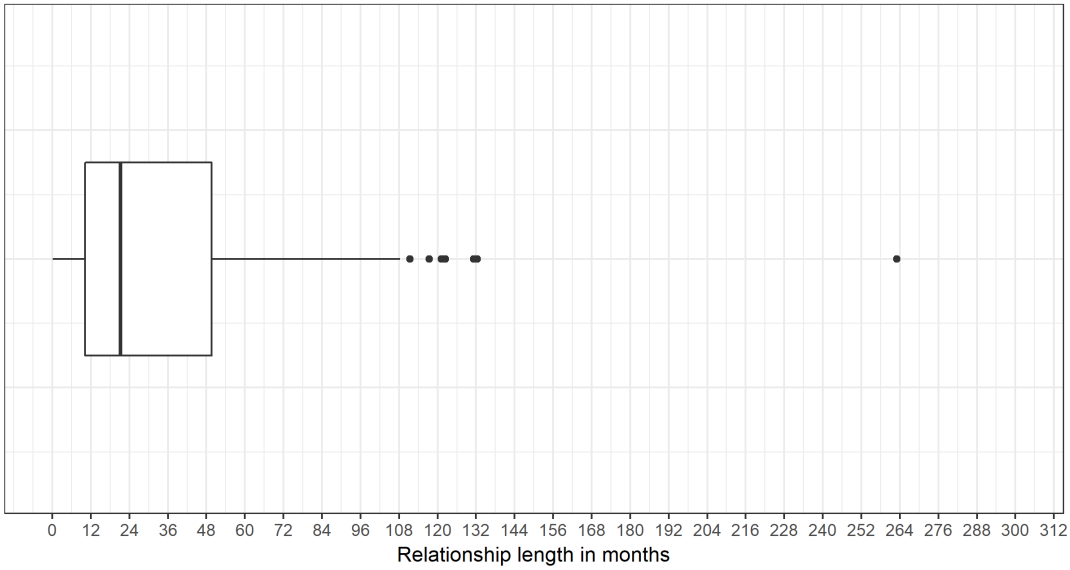


Note: The boxplot shows the distribution of couples over the normalised consistncy score. It is calculated by dividing the number of points a couple scored in answering the questions documented in Table 4.6 over the maximum number of points.

[51] There is no obvious choice for the number of points an open question like this should receive. I chose to assign the maximum number that can be scored in a closed question.

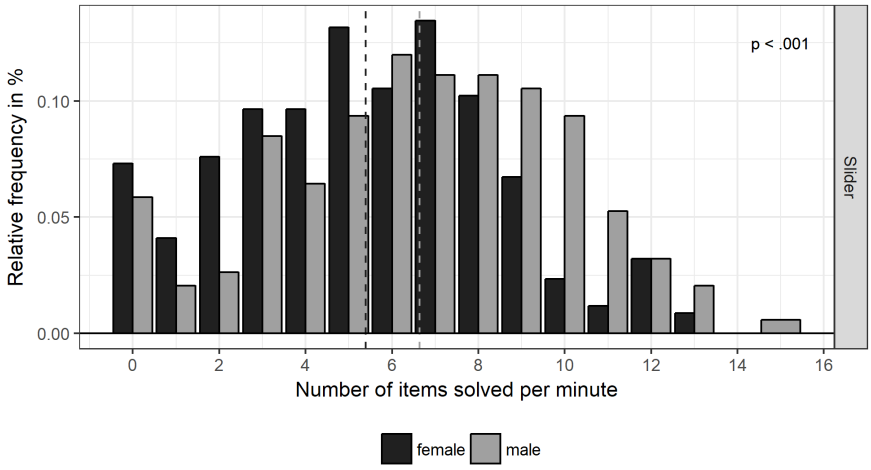
4.8 Appendix C: Additional results

Figure 4-7: Relationship duration



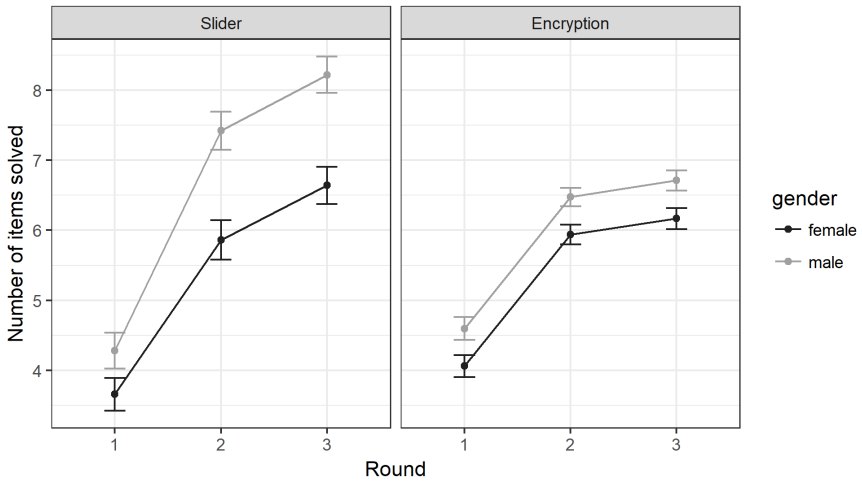
Note: Relationship duration is calulated as the number of days that have elapsed between the self-reported starting date of the relationship and the date on which the session took place.

Figure 4-8: Stage 1 Slider task productivity by gender



Note: Performance data pooled from 3 rounds in stage 1. Dashed lines show means, p-values are obtained from a Kolmogorov-Smirnov test for the equality of distributions.

Figure 4-9: Stage 1 Encryption and Slider task productivity by round and gender



Note: Each round took one minute. Round 1 was an unpaid practise round. In Round 2 and 3 participants received a piece rate of 10 and 20ECU, respectively.

Table 4.7: Proportion of income maximising choices, no information on partner productivity

	Choice: Income maximising			
	(1)	(2)	(3)	(4)
Intercept	1.001***	0.996***	0.981***	1.015***
	(0.044)	(0.043)	(0.061)	(0.059)
PA	−0.596***			
	(0.050)			
$PA \times Stranger$	−0.069*			
	(0.037)			
PA^m		−0.551***	−0.489***	−0.618***
		(0.054)	(0.077)	(0.075)
$PA^m \times Stranger$		−0.057	−0.038	−0.075
		(0.048)	(0.069)	(0.066)
PA^f		−0.676***	−0.709***	−0.641***
		(0.064)	(0.092)	(0.088)
$PA^f \times Stranger$		−0.059	0.001	−0.123
		(0.063)	(0.090)	(0.088)
Productivity controls	N	N	N	N
Information	N	N	N	N
Sample	All	All	Men	Women
Observations	492	492	246	246

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are reported in parantheses. Coefficients are obtained from the model described in Section 4.4, equation 4.9.

Table 4.8: Proportion of income maximising choices, no information on partner productivity, including controls

	Choice: Income maximising			
	(1)	(2)	(3)	(4)
Intercept	0.995^{***}	0.993^{***}	0.982^{***}	1.011^{***}
	(0.042)	(0.041)	(0.061)	(0.057)
PA	-0.795^{***}			
	(0.079)			
$PA \times Stranger$	-0.085^{**}			
	(0.037)			
PA^m		-0.743^{***}	-0.806^{***}	-0.692^{***}
		(0.082)	(0.117)	(0.112)
$PA^m \times Stranger$		-0.092^{*}	-0.061	-0.124^{*}
		(0.048)	(0.070)	(0.064)
PA^f		-0.865^{***}	-1.017^{***}	-0.716^{***}
		(0.088)	(0.127)	(0.119)
$PA^f \times Stranger$		-0.046	0.007	-0.109
		(0.063)	(0.090)	(0.085)
Productivity controls	Y	Y	Y	Y
Information	N	N	N	N
Sample	All	All	Men	Women
Observations	492	492	246	246

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are reported in parantheses. Coefficients are obtained from the model described in Section 4.4, equation 4.9.

Table 4.9: Proportion of income maximising choices *with* information on partner productivity, including controls

	Choice: Income maximising			
	(1)	(2)	(3)	(4)
Intercept	0.998***	0.998***	0.993***	1.005***
	(0.020)	(0.020)	(0.029)	(0.029)
PA	−0.546***			
	(0.040)			
$PA \times Stranger$	0.057*			
	(0.034)			
$PA \times Computer$	0.091***			
	(0.026)			
PA^m		−0.542***	−0.510***	−0.607***
		(0.044)	(0.062)	(0.062)
$PA^m \times Stranger$		0.060	−0.013	0.128**
		(0.043)	(0.061)	(0.060)
$PA^m \times Computer$		0.056*	0.053	0.075*
		(0.032)	(0.047)	(0.045)
PA^f		−0.563***	−0.487***	−0.668***
		(0.053)	(0.075)	(0.074)
$PA^f \times Stranger$		0.057	−0.100	0.216***
		(0.056)	(0.079)	(0.078)
$PA^f \times Computer$		0.139***	−0.004	0.308***
		(0.044)	(0.062)	(0.063)
Productivity controls	Y	Y	Y	Y
Information	Y	Y	Y	Y
Sample	All	All	Men	Women
Observations	1,722	1,722	861	861

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are reported in parantheses. Coefficients are obtained from the model described in Section 4.4, equation 4.9.

Table 4.10: Proportion of income maximising choices *with* information on partner productivity, including controls, excluding 'wrong' breadwinner choices

	Choice: Income maximising			
	(1)	(2)	(3)	(4)
Intercept	0.999***	0.999***	0.991***	1.006***
	(0.018)	(0.018)	(0.025)	(0.026)
PA	−0.485***			
	(0.037)			
$PA \times Stranger$	0.056*			
	(0.031)			
$PA \times Computer$	0.128***			
	(0.024)			
PA^m		−0.467***	−0.419***	−0.542***
		(0.040)	(0.055)	(0.059)
$PA^m \times Stranger$		0.062	0.017	0.104*
		(0.040)	(0.054)	(0.057)
$PA^m \times Computer$		0.100***	0.104**	0.113***
		(0.030)	(0.041)	(0.043)
PA^f		−0.519***	−0.447***	−0.618***
		(0.048)	(0.065)	(0.069)
$PA^f \times Stranger$		0.060	−0.077	0.199***
		(0.050)	(0.068)	(0.072)
$PA^f \times Computer$		0.174***	0.073	0.294***
		(0.040)	(0.053)	(0.059)
Productivity controls	Y	Y	Y	Y
Information	Y	Y	Y	Y
Sample	All	All	Men	Women
Observations	1,627	1,627	808	819

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are reported in parantheses. Coefficients are obtained from the model described in Section 4.4, equation 4.9.

Table 4.11: Proportion of income maximising choices *with* information on partner productivity, including controls, excluding 'wrong' dual earner choices

	Choice: Income maximising			
	(1)	(2)	(3)	(4)
Intercept	1.000***	1.000***	1.002***	0.999***
	(0.014)	(0.014)	(0.021)	(0.018)
PA	−0.139***			
	(0.030)			
$PA \times Stranger$	0.003			
	(0.025)			
$PA \times Computer$	−0.033*			
	(0.020)			
PA^m		−0.164***	−0.169***	−0.178***
		(0.032)	(0.047)	(0.044)
$PA^m \times Stranger$		0.004	−0.050	0.059
		(0.032)	(0.047)	(0.042)
$PA^m \times Computer$		−0.039	−0.051	−0.022
		(0.024)	(0.036)	(0.032)
PA^f		−0.097**	−0.090	−0.131**
		(0.041)	(0.059)	(0.056)
$PA^f \times Stranger$		−0.014	−0.067	0.045
		(0.044)	(0.065)	(0.058)
$PA^f \times Computer$		−0.045	−0.108**	0.041
		(0.035)	(0.049)	(0.049)
Productivity controls	Y	Y	Y	Y
Information	Y	Y	Y	Y
Sample	All	All	Men	Women
Observations	1,509	1,509	762	747

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are reported in parantheses. Coefficients are obtained from the model described in Section 4.4, equation 4.9.

Table 4.12: Stage 3 wage earner effort supply, including controls

	Effort		
	(1)	(2)	(3)
Intercept	12.080*** (1.487)	11.239*** (1.754)	14.913*** (2.192)
Stranger	1.837 (1.130)	1.747* (1.018)	−0.560 (1.233)
Computer	1.925 (1.281)	2.496** (1.222)	−0.840 (1.613)
Partner_mp	3.679*** (1.377)	3.671*** (1.295)	0.284 (1.592)
SE	2.849** (1.159)	2.173* (1.118)	−0.269 (1.530)
Stranger × *Partner_mp*	−2.495 (1.799)	−2.310 (1.630)	−0.230 (2.081)
Computer × *Partner_mp*	−2.563 (2.675)	−2.333 (2.452)	−1.622 (2.857)
Stranger × *SE*	−3.260** (1.442)	−3.039** (1.301)	0.968 (1.839)
Computer × *SE*	−1.759 (1.575)	−1.211 (1.442)	2.356 (2.079)
Partner_mp × *SE*	0.964 (2.799)	1.128 (2.530)	5.465** (2.355)
Stranger × *Partner_mp* × *SE*	−1.740 (4.333)	−1.657 (3.920)	−2.749 (3.992)
Computer × *Partner_mp* × *SE*	0.102 (4.746)	−0.592 (4.302)	2.123 (4.251)
fem	1.499 (1.121)		
fem × *Stranger*	−2.315 (1.587)		
fem × *Computer*	−2.516 (1.869)		
fem × *Partner_mp*	−3.858** (1.874)		
fem × *SE*	−3.173* (1.697)		
fem × *Stranger* × *Partner_mp*	1.831 (2.601)		
fem × *Computer* × *Partner_mp*	0.743 (3.658)		
fem × *Stranger* × *SE*	4.146* (2.192)		
fem × *Computer* × *SE*	3.754 (2.434)		
fem × *Partner_mp* × *SE*	4.326 (3.528)		
fem × *Stranger* × *Partner_mp* × *SE*	−0.352 (5.640)		
fem × *Computer* × *Partner_mp* × *SE*	3.992 (6.026)		
Controls	Y	Y	Y
Sample	All	Men	Women
Observations	462	249	213
R^2	0.538	0.567	0.506

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01, standard errors are clustered at the couple level and reported in parantheses. Coefficients are obtained from the model described in Section 4.4, equation 4.11. The set of controls comprises (i) the number of times the male breadwinner option was chosen with partner p, and (ii) the number of times the female breadwinner option was chosen, (iii) i's stage 1 productivity, and (iv) her partner's stage 1 productivity, (v) the order of play (stranger/partner), (vi) the total number of times i was selected as performer of the paid task in stage 3.

Chapter 5

Choosing between career & family – Gender roles as coordination device[1]

Abstract

This chapter investigates the role of gender norms as a coordination device and efficiency enhancer. 192 subjects, real heterosexual couples, play a classic *battle of the sexes* game once with their partner and once with a randomly matched stranger. The strategies were framed neutrally in the control group (A vs. B) and as a family specialisation game in the treatment group ($Career$ vs. $Family$). The probability that couples coordinate on the traditional gender role equilibrium increases in the *Specialisation* treatment, yet the overall probability to coordinate, and thus efficiency, does not improve much. This is driven by heterogeneous responses to norms: Compared to the control group, women opt for $Career$ at a significantly lower rate in the *Specialisation* treatment. Men, however, are only more likely to opt for $Career$ when they play with a stranger, but not with their real partner. I present evidence that some men may have motive to signal progressive gender norms when playing with their partner.

[1]Valuable comments by Ralph Bayer, Miriam Beblo, Elisabeth Bublitz, Evelyn Korn, Eva Markowsky, Gerd Mühlheußer, Kristin Paetz, Elizabeth Peters, Arne Pieters, Melanie Schröder, as well as participants of the 2018 ASSA meetings and the 2018 Leuphana Workshop on Microeconomics are gratefully acknowledged. I thank the UHH WiSo Graduate School and Miriam Beblo for providing funds in support of this project.

5.1 Introduction

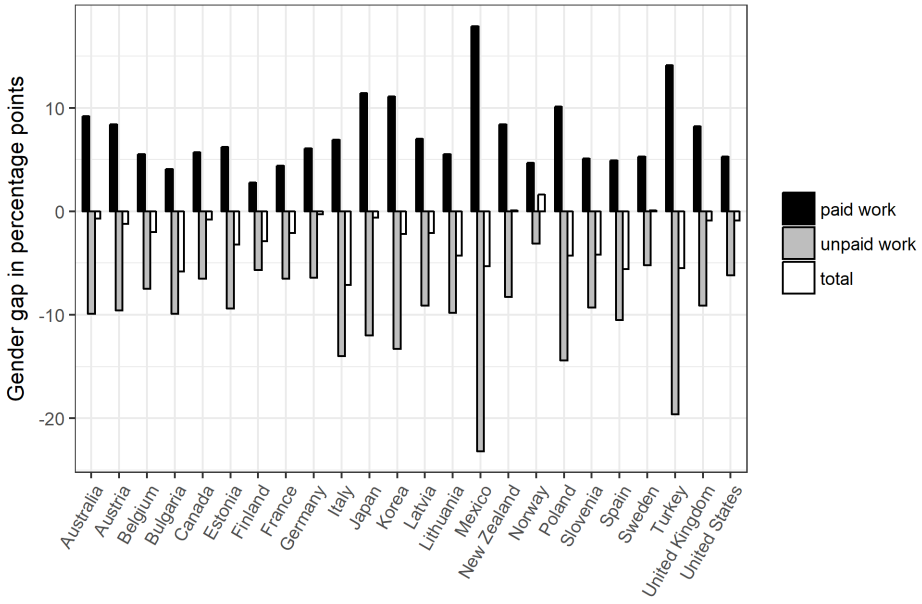
After a period of convergence starting in the beginning of the last century, gender gaps in labour force participation have come to stagnate in most industrialised countries (Goldin, 2014). This appears to be closely linked to the gendered division of labour in families, which usually requires women to spend less time in paid market work and more time engaged in unpaid house and care work than men.[2] Even in countries where participation rates are closer to parity, childbirth pushes women, but not men, into jobs that are more compatible with care work, thus paving the way for substantial career gaps (Kleven et al., 2018; Blau and Kahn, 2017). As a result, gender inequality amplifies over the life cycle on many dimensions: Women, and mothers in particular, receive lower lifetime earnings (Adda et al., 2017; Goldin et al., 2017), savings, and pension entitlements (OECD, 2017). Economic dependence on male partners makes women more vulnerable to economic hardship in case of divorce or death of a partner (OECD, 2017a), as well as to domestic violence (Bobonis et al., 2013; Anderberg et al., 2016).

Figure 5-1 shows that the gendered division of domestic and market labour is common practice in industrialised countries across the world. Using data provided by the OECD, the graph displays the gender gaps in the shares of daily time spent in paid and unpaid work in each member state. Two aspects are noteworthy: First, without exception, gender gaps in paid work are positive, and negative in unpaid work. That is, on average, men spend a larger proportion of their total daily time working for pay than women, while the reverse is true for unpaid work. Second, the gap in total work, i.e., the time spent in paid and unpaid work combined, are

[2]Homosexual households tend to specialise less (Black et al., 2007), and there may be several explanations for this. The gains from specialisation are certainly lower in childless households, and same-sex couples are more likely to fall into this category. Moreover, restricted access to legal marriage may inhibit the potential to enter into binding contracts, thereby making specialisation less appealing, especially to the individual taking on the homemaker role. Finally, heteronormative gender roles may be less powerful in homosexual partnerships. This chapter focusses on heterosexual couples.

also negative in most countries, i.e., women spend a larger fraction of their day working than men.[3]

Figure 5-1: Gender gaps in percentage of total time spent in paid and unpaid work



Note: Own calculations based on OECD family database (2016) data, a compilation of time use studies from each country containing information on men's and women's (self-reported) time spent on work activities as a percentage share of total daily time. Gender gaps are calculated as the difference in male vs. female country averages. The category paid work is taken as is from the data, while the category unpaid work is formed by aggregating unpaid and care work (excluding personal care). The category total work is formed by aggregating paidand unpaid work.

Economists have considered different explanations for the gendered division of labour. Perhaps the most pervasive idea stems from the work of Gary Becker, 1973 and Reuben Gronau, 1973a,b. Applying the Ricardian principle of comparative advantage to household production, these models show that, in the presence of relative productivity differences, specialisation allows partners to generate efficiency gains. While, at the time, there was little controversy in assuming women

[3]Note that a caveat of the data is that the fractions of time spent in a certain activity are self-reported, so gender differences in reporting behaviour cannot be precluded. The residual categories are the fractions of time spent on: a) personal care; b) leisure; or c) unspecified.

were generally less productive in the labour market than men,[4] the main result of these models is gender-neutral. Thus, they predict a reversal of traditional specialisation—i.e., a female breadwinner and male homemaker—in households where the female partner holds a *relative* productivity advantage in market work.[5]

Almost half a century later, female breadwinners remain the exception in heterosexual households, in spite of women's rapidly increasing labour market productivity since that time. This led economists to question whether specialisation decisions are sufficiently described by the gender-neutral household production models proposed by Becker (1973) and Gronau (1973a), or whether gender norms factor in. Specifically, they might reduce productive efficiency in households where the female spouse has a comparative advantage for market work because they prescribe choosing the male as the breadwinner. Akerlof and Kranton (2000), in their seminal paper on identity economics, provide the tools to incorporate a penalty for 'norm violation' into a formal economic analysis, and discuss its implications for the household division of labour, which will be asymmetric in the presence of a social norm prescribing that 'a man should not do women's work,' even for symmetric productivities. In this spirit, Cudeville and Recoules (2015) present a theoretical model to analyse how gender norms, combined with couples' desire for social conformity, distorts male and female labour supplies when women have a comparative advantage for market work. The papers by Bertrand et al. (2015) and Görges (2015) provide empirical support for this notion.

[4]Both Becker (1973) and Gronau (1973a) argued that women's lower productivity in the labour market was apparent from their lower wages. While Becker (1985) later explored the possibility that women's higher responsibility for home production causes the gender wage gap, he maintained the assumption of gender differences in relative productivity in his later accounts of household production (Becker, 1991). Notably, though, he then attributed them primarily to women's biological advantage in childrearing rather than gender differences in labour market productivity.

[5]As was discussed in greater detail in the preceding Chapter 4, the destinction between absolute and relative advantage is crucial for the household specialisation problem. If women's productivity advantage in the home is sufficiently large, traditional gender roles may not reverse even when they hold an absolute advantage in market production.

In contrast to this recent focus on gender norms as a potential distortion of productive efficiency, other economists have used game theory to explore their potential as efficiency enhancers. Lundberg and Pollak (1993) suggest a "separate spheres" equilibrium, in which spouses specialise according to gender roles that serve as a focal point. Engineer and Welling (1999) and Hadfield (1999) propose coordination models in which marriages are more efficient for matches with comparative advantages. As a result, individuals have an incentive to invest in gender-specific skills prior to marriage in order to coordinate with a future partner. They show that, while a mixed gender equilibrium, it is less stable and less efficient than a gendered one. Baker and Jacobsen (2007) use a similar set-up to show that gender norms may mitigate both coordination and marital hold-up problems. This literature thus suggests a theoretical argument for why gender norms could be socially desirable, as they may facilitate coordination and improve overall efficiency.[6]

To the best of my knowledge, the idea that gender norms can improve productive efficiency among non-cooperative spouses has not been tested empirically. To address this gap, this chapter uses a lab experiment to study whether gender norms provide a social cue—a *focal point* (Schelling, 1960)—that couples use as a coordination device. A major advantage of a lab experiment is that it allows one to study the direct impact of the presence of gender norms, imposing identical rates of productivity and preferences for men and women. This is difficult to achieve with survey data, since gender norms may likely also have an indirect effect on these productivity and preference variables. Recent experimental studies of focal points provide mixed evidence: Some find that their efficiency-enhancing potential is large, both in tacit (Isoni et al., 2013) and explicit bargaining (Isoni et al., 2014), while others emphasise its limits (Crawford et al., 2008). The fact that these differing results are obtained from different experimental settings, in which the salience of the focal point varied, indicates that the conclusions may not be

[6]Baker and Jacobsen (2007), however, also stress that these gains may not be Pareto-improving for both genders.

readily transferred to a different context, such as intra-household labour devision. Despite their relavance to that specific context, this is the first paper to study how focal points affect *real* heterosexual couples playing a classic *battle of the sexes*, i.e., a symmetric coordination game where the two pure strategy Nash equilibria result in unequal payoffs that favour either the male or female partner. Thus, this particular study differs on two important dimensions: First, I observe subjects' choices when playing with their partner and with a randomly matched stranger, thus gaining insights that go beyond what can be extracted from a standard student subject pool. Second, I vary the presence of gender norms by presenting the two strategies with neutral labels (A vs. B) or labeled as a family specialisation decision ($Career$ vs. $Family$).

The findings of this study are surprising in three respects: First, overall coordination rates improve only by a small margin in the *Specialisation* treatment, while payoff inequality increases dramatically to the benefit of male players. That is, the probability that couples coordinate on the traditional gender role equilibrium increases, yet efficiency does not increase by much. Second, the effect is equally present among real couples and pairs of randomly matched strangers, for which post-experimental redistribution of earnings cannot occur. Finally, an investigation of individual choices by gender reveals an unexpected pattern: Compared to the $Neutral$ control group, women opt for $Career$ at a significantly lower rate in the *Specialisation* treatment, regardless of familiarity with the partner. Men, however, are only more likely to opt for $Career$ when they play with a stranger, but not with their real partner. I investigate two different mechanisms that might be driving this. Beliefs regarding partners' $Career$ choice do not seem to provide a motive. Instead, I find suggestive evidence for "marriage market signaling" in the spirit of Bursztyn et al. (2017); namely, that some men might want to signal progressive gender norms to their potential long-term partners (but not to an anonymous randomly matched woman). The chapter is organised as follows: Section 5.2 lays out the theoretical background and describes the game, while details

on the experimental procedure are provided in Section 5.3. The main results are supplied in Section 5.4, and Section 5.5 presents analyses regarding the mechanism. Section 5.6 offers a discussion and concludes the chapter.

5.2 Theoretical background & the game

Consider a non-cooperative household consisting of a male partner, m, and a female, f. Individual utilities are given by $U_i(y, c)$, $i = \{f, m\}$, where y is monetary income, which can only be acquired by supplying labour to the market, and c is the quality of family life, including children, which can only be produced in the home. Each household member is endowed with one unit of time and makes a binary choice regarding its investment. They can either invest all their time into their *Career*, i.e. engage in market work w and produce y. Alternatively, they can invest all their time into *Family* work, i.e., engage in home production h and produce c. Both y and c are public goods, i.e., consumed by both partners once produced.[7] Assume further that preferences are such that both arguments of the utility function are essential in the sense of Hart and Moore (1990), i.e., $U_i(y, \bullet) = U_i(\bullet, c) = 0$. Denote by $U_i^{a_i,a_j}$ the utility of partner i engaging in activity a_i when her partner engages in a_j, where $a_{i,j} = \{w, h\}$. This implies that $U_i^{h,h} = U_i^{w,w} = 0$.

Now, if partners have identical preferences over y and c, and are equally productive in both lines of work w and h, this further implies $U_m^{w,h} + U_f^{h,w} = U_m^{h,w} + U_f^{w,h}$. Finally, assuming that individual payoff is higher when engaging in market work rather than in family work, but maintaining that there are no intrinsic differences

[7] Commodities that are produced in the home are typically assumed to have public good character, such as a clean house or child-wellbeing, the example used here. The main argument is that parents derive utility from their children and value their wellbeing, oftentimes even in the case of divorce Browning et al. (2014). Market income, on the other hand, is commonly treated as a private good in the literature. However, the assumption that income is a public good has been justified by Beblo and Robledo (2008) on the grounds that the bulk of it is spent on public goods (e.g. housing) and that most families use joint bank accounts. One may also argue that social norms dictate the sharing of income between spouses, at least in some cultures.

in preferences between the sexes, we have $U_m^{h,w} = U_f^{h,w} < U_m^{w,h} = U_f^{w,h}$. This last assumption can be justified in a number of ways. An intuitive reason is that there may be a social norm of sharing income with your spouse that makes labour earnings a public good, but that the spouse who acquires these earnings may either have private information about the amount, or some discretionary power over it, which allows her to determine a split in her favour. Both phenomena—hiding of income or windfalls from spouses and exercising control over individually obtained income—are well-documented in the experimental family economics literature (Ashraf, 2009; Iversen et al., 2011; Mani, Mani; Castilla and Walker, 2013; Kebede et al., 2013; Ambler, 2015; Hoel, 2015).[8]

The simple model is summarised in a 2x2 matrix in Table 5.1 below. Each player can choose between two actions (labeled *Career* and *Family* in the *Specialisation* treatment, *A* and *B* in the *Neutral* control group). For a given action chosen by j, i maximises her own pay-off by *not* matching j's action. Subjects were presented with this standard coordination game in the spirit of the classic *battle of the sexes* and made a one-shot decision, once matched with their real partner and once with a stranger.[9] Payoffs were set at $a = 200$ ECU and $b = 100$ ECU.

Applying standard game theory, there are two equilibria in pure strategies in this game: (*Career*, *Family*) and (*Family*, *Career*). In the neutrally framed version of the game played in the control group, the mixed strategy equilibrium involves playing *A* (*Career*) with probability $\frac{a}{a+b}$ and *B* (*Family*) with $\frac{b}{a+b}$, i.e.

[8] Another possibilty is that the differences in payoffs reflect the consequences of the individual time investment for a future period that is not explicitly modeled. Imagine, for example, that in period 2, the children have grown up and left the household, but are still a public good to their parents in the sense that both parents have access to them and derive utility from them regardless of the parents' marital state. Both spouses will now invest all their time in market work, but levels of wage productivity have diverged because j has been absent from the labour market in period 1. While married, social norms may dictate pooling the labour incomes and guarantee an equal split for both spouses. By divorcing j, however, i can restrict j's access to her labour income. Under certain conditions, i would prefer to divorce and consume the same amount of c as j, but a higher amount of income y.

[9] While the matrix representation is familiar to economists, the experimental instructions used a more intuitive illustration to convey the decision structure to subjects from different backgrounds. See Appendix 5.7 for details.

Table 5.1: Battle of Sexes, $a > b > 0$

		Player 2	
		A / Career	*B / Family*
Player 1	*A / Career*	$0, 0$	a, b
	B / Family	b, a	$0, 0$

Note: Strategies were labeled A and B in the *Neutral* control, and *Career* and *Family* in the *Specialisation* treatment.

$\frac{2}{3}$ and $\frac{1}{3}$, respectively, given the payoffs chosen in the experimental implementation. Expected payoffs are $\frac{ab}{a+b}$, i.e. nearly 67 ECU, and the probability to coordinate is just below 45%. When the available actions are labeled *Family* and *Career*, however, coordination rates should improve according to the *focal point effect* (Schelling, 1960), which should induce rational players to play the equilibrium that is culturally dominant.[10]

Analysing the household specialisation problem as a non-cooperative bargaining game is particularly useful for two reasons: First, the case in which both partners are equally productive becomes increasingly relevant in light of women's educational achievements in recent years. Thus, young couples are presumably more likely than older cohorts to find themselves in a situation where no 'objective' criterion can readily determine the primary producer of the household good. Therefore, the *battle of the sexes* game presented here may present a use-

[10]In light of Holm's (2000) findings, it also seems plausible that, even in the *Neutral* treatment, subjects coordinate more often on the equilibrium in which the male player receives the higher payoff. Arguing that gender itself can serve as a focal point, he investigates subjects' strategies in a neutrally framed *battle of the sexes*, providing information on the opponent's gender. In three different studies with Swedish and American student subjects, he consistently finds that subjects behave significantly more 'hawkish' when playing against a female player. Nonetheless, it seems plausible that the labeling of the strategies as *Family* and *Career* provides a more salient focal point, which should increase coordination rates in the treatment vis-a-vis the control group.

ful, albeit simplified, approximation to an increasingly relevant real-world decision problem. Second, even though most family economic models assume cooperation among household members and thus predict productive efficiency (Manser and Brown, 1980; McElroy and Horney, 1981; Chiappori, 1988), others suggest that non-cooperative behaviour should at least be considered as the threat point (Lundberg and Pollak, 1994; Konrad and Lommerud, 2000). Furthermore, mounting empirical literature on experiments with family members by and large suggests non-cooperative behaviour to be common (Peters et al., 2004; Ashraf, 2009; Iversen et al., 2011; Mani, Mani; Castilla and Walker, 2013; Kebede et al., 2013; Ashraf et al., 2014; Cochard et al., 2014; Ambler, 2015; Beblo et al., 2015; Hoel, 2015; Beblo and Beninger, 2017).

In light of these considerations, it seems reasonable to study non-cooperative models of family decision making, e.g., as Konrad and Lommerud (1995) and Chen and Woolley (2001) did theoretically by analysing Cournot models of family public good provision. Konrad and Lommerud (1995) study a non-cooperative application of the household specialisation problem and, perhaps surprisingly, reproduce the main result of Becker (1973) and Gronau (1973a). They show that the spouse who is more productive in the home allocates more time to producing the household public good than their partner.[11] Contrary to the Becker and Gronau model, however, their model predicts that time investments in producing the household public good will be inefficiently low (Konrad and Lommerud, 1995).

Raising the public good provision (closer) to its efficient level may be achieved in two ways. First, spouses may enter into a binding contract specifying an enforceable transfer from the spouse who does not contribute to the public good to the one who does, such that it raises the overall contribution level. If this is possible, the household's optimisation problem becomes essentially equivalent to the Becker (1973) and Gronau (1973a) model, in which marriage is implicitly assumed

[11] Unlike the models of Becker (1973) and Gronau (1973a), however, they show that absolute advantage determine contribution levels and argue that, given the gender wage gap, this makes women systematically worse off than men (Konrad and Lommerud, 1995).

to provide such a contract, and will thus yield efficient results. One concern is of course that the contract provided by marriage may not be complete in the sense that it covers all relevant contingencies (Hart and Moore, 1988), a problem that has become more apparent in light of rising divorce rates and weakened alimony laws in many countries (e.g., for the US, see Stevenson and Wolfers, 2007). The contractual incompleteness of marriage may impede couples' ability to cooperate and lead to inefficient public good investments (e.g., lower fertility; see Fahn et al., 2016). Social norms, combined with individuals' desire to adhere to behavioural prescriptions, may provide an alternative way to nevertheless raise public good provision above the non-cooperative benchmark by prescribing that a specific partner must produce the socially efficient amount (or suffer utility losses from violating the norm).

Against this background, the results of this experiment can be informative in several ways: First, since by design no gender differences exist in productivity or preferences, we can assess the question of whether gender norms, independent of other considerations, influence specialisation decisions by comparing the choices of the *Neutral* control group to those of the *Specialisation* treatment. If they do, the second important question is whether they actually help improve coordination rates and thus productive efficiency. Increasing efficiency by steering individuals toward the traditional gender equilibrium comes at the cost of female players, and comparing women's losses in the *Specialisation* treatment to the control, and across pairings with their real partner and a stranger allows us to assess how costly these norms are to women.

Finally, comparing the change in coordination rates across real partners and strangers might reveal whether individual norms can superimpose societal norms. For example, if individuals hold progressive gender norms, they might play $(a_m, a_f) = (Family, Career)$ with their partners but rationally expect a randomly drawn stranger from a representative distribution of gender norms to play $(a_m, a_f) = (Career, Family)$. Thus, we might see similar coordination rates for couples and

strangers in the treatment, but the proportion of real couples playing the progressive equilibrium might be larger. The reverse phenomenon might occur when subjects expect that gender norms in the student population are on average more progressive than their own.[12] I use two different elicitation methods to capture individual beliefs about their partner's individual and the 'general' norms to investigate this. Detailed descriptions of the procedures are provided in the following section.

5.3 Experimental procedure

The experiment was programmed in ztree (Fischbacher, 2007) and conducted in the economics lab at the University of Hamburg in 2016 and 2017. A total of 192 subjects were recruited via hroot (Bock et al., 2014) from a regular student subject pool. The *battle of the sexes* game analysed in this chapter was part of a larger experiment and took about 10 minutes in sessions lasting 2.5 hours on average. Average payouts for the experiment were €27.43 (approximately $34 at the time

[12]If subjects are inequality-averse, they may prefer an equal payoff of 0 over an unequal distribution, and thus try to prevent coordination. When matched with an inequality-neutral player, this type of strategy could be more successful in the *Specialisation* treatment group, since *naive* inequality-neutral opponents will select a strategy expecting to reach the culturally dominant equilibrium. Accordingly, this should reduce coordination rates in the treatment compared to the control group, or at least diminish the efficiency gains obtained in the treatment, in the presence of different types (inequality-averse and neutral). The presence of inequality-averse players could also affect the coordination rates realised by real partners versus strangers: Inequality-averse players may be willing to coordinate with their real partners if they expect a redistribution of income after the experiment. They would, however, try to prevent coordination with a stranger because anonymity prohibits any form of ex-post redistribution between them. Since post-experimental compensation can only be expected in cooperative couples, inequality-averse individuals in non-cooperative couples should avoid coordination regardless of the partner they are paired with. However, since subjects would have to exhibit very high degrees of advantageous and disadvantageous inequality aversion, given the payoffs chosen in this experiment, I do not consider it as a motive. Using the Fehr and Schmidt (1999) model as an example, the parameter for (dis)advantageous inequality aversion would have to be equal to 2 (1), which is very large. In their large-scale experiments, Charness and Rabin (2002) have shown that the majority of individuals are most strongly concerned with raising social welfare.

of writing this chapter).[13] A total of 9 sessions were conducted, with a median of 20 subjects per session; the largest (smallest) session had 28 (12) subjects. Four sessions (80 participants) were assigned to the treatment (*Specialisation*), and the remaining five (112 participants) formed the control (*Neutral*). All subjects played the game with both their real partner and a randomly matched stranger, where partner gender was held constant across pairings.[14] Thus, the data features between-subject variation in the presence of a gender norm focal point and within-subject variation in the pairing.

The invitation email instructed subjects to pre-register their partner and bring them to the experimental session. Being married or cohabitating with their partner was not required, nor were partners required to be enrolled as students. Upon arrival in the lab, participants were reminded that they needed to be true couples in a relationship, not merely study partners or housemates, in order to participate in the experiment. They were asked to leave if this was not the case; however, no one left the experiment. Additional ex-post checks to mitigate concerns regarding potential "fake couples" are provided in Appendix 5.8.

Instructions for the game explicated the payoff consequences of the two available actions for both possible scenarios (partner choosing *Career* or *Family*). They further informed subjects that they would play once with their partner and once with a randomly matched stranger (same sex as partner), and that they would have no knowledge of their partner's chosen action when making their decision, but would find out about each other's respective choices once the payouts were revealed.[15] After having read the instructions, subjects were paired with their first partner (real or stranger) and entered their choice. Next, they were asked to predict their partner's decision. Upon completing this step, but before finding out the result, they were paired with their second partner (real or stranger) to re-

[13]Throughout the experiment, all money amounts were expressed in ECU (Experimental Currency Units), which were later converted to Euros at an exchange rate of 1 = .01.

[14]Order effects are unlikely because the result of each interaction was not revealed until both had been completed.

[15]A translated version of the instructions can be found in Appendix 5.7.

peat the previous steps. Once completed, subjects were confronted with four more prediction tasks: How many women (men) had chosen *Career* when playing with their real partner (stranger)? All prediction tasks were incentivised.[16] Finally, the outcomes of both games and the corresponding payoffs were revealed.

Descriptive statistics are reported in Table 5.2.[17] All sessions were gender balanced by design. On average, participants in the *Neutral* control were about 25 years old, and 27 years old in the *Specialisation* treatment. Because the difference is significant, age will be included as a control in subsequent analyses.[18] Union characteristics are measures that describe subjects' relationships with their partner. Most were very similar across treatments: The average union duration (time spent in the relationship with their partner) was three years, almost a third of the couples were cohabitating, and participants reported high satisfaction levels with their relationship on average. The treatment and control groups differ with respect to the proportion of married couples: 7% of the participants in the *Neutral* control were married, while no couples in the *Specialisation* treatment were married. Consequently, marital status will also be included as a control variable in subsequent analyses. Most subjects were currently enrolled in some form of educational institution, primarily university. The treatment and control samples do not differ statistically with respect to enrollment status or the level of study. Furthermore,

[16]Throughout the experiment, subjects were asked to make several predictions and were informed that two-thirds of the predictions would be randomly selected and paid out. They received 50ECU for correctly predicting their partners' choice between *Career* vs. *Family*, and 100ECU for the total number of men (women) choosing *Career* in their session.

[17]The data analyses and the presentation of results in this chapter were prepared using the *R programming environment for statistical computing* (R Core Team, 2018), version 3.5.0., and several add-on packages: *broom* (Robinson, 2018), *car* (Fox and Weisberg, 2011), *lmtest* (Zeileis and Hothorn, 2002), *multiwaycov* (Graham et al., 2016), *plm* (Millo, 2017), *stargazer* (Hlavac, 2018), *sandwich* (Zeileis, 2006), *tidyverse* (Wickham, 2017), *xlsx* (Dragulescu, 2014), and *xtable* (Dahl, 2016). For calculating cluster-robust confidence intervals, I use a function written by Gubler (2014).

[18]The main concern here is that potential differences between groups may be driven by differences in age rather than the treatment. Since there is no obvious reason why the response to treatment should vary with age, interaction effects for age and the other control variables mentioned in this paragraph are not included.

the two groups are very similar with respect to labour market activity. About a third of subjects are not active at all, and the rest are either unemployed or employed (full-time, part-time, or irregularly). Finally, the samples differ with respect to the proportion of subjects who ever participated in a lab experiment; thus, this indicator variable will be included in the subsequent analyses as well.

Table 5.2: Summary statistics by treatment

	Neutral (N)		*Specialisation* (S)		Difference ($N - S$)	
variable	N	Mean	N	Mean	Mean	p-value
Demographics						
Female	112	0.50 (.)	80	0.50 (.)	0.00	1.000
Age*	112	24.75 (4.36)	80	27.09 (5.5)	-2.34	0.002
Union charateristics						
Union duration*	112	3.00 (2.98)	80	3.00 (4.26)	0.00	0.995
Cohabitating	112	0.27 (.)	80	0.28 (.)	-0.01	1.000
Married	112	0.07 (.)	80	0.00 (.)	0.07	0.038
Satisfaction with union*	112	8.83 (1.45)	80	8.53 (1.95)	0.31	0.238
Currently in education						
Not in education	112	0.12 (.)	80	0.16 (.)	-0.04	0.599
In school	98	0.03 (.)	67	0.01 (.)	0.02	0.898
Undergraduate student	95	0.61 (.)	66	0.56 (.)	0.05	0.638
Postgraduate student	95	0.23 (.)	66	0.30 (.)	-0.07	0.405
Other student	95	0.16 (.)	66	0.14 (.)	0.02	0.879
Labour market activty						
Not active in the labour market	112	0.29 (.)	80	0.31 (.)	-0.02	0.915
Full time employed	112	0.16 (.)	80	0.20 (.)	-0.04	0.609
Part-time employed	112	0.24 (.)	80	0.20 (.)	0.04	0.619
Irregularly employed	112	0.30 (.)	80	0.29 (.)	0.02	0.936
Unemployed	112	0.04 (.)	80	0.05 (.)	-0.01	0.903
Experience with economics and experiments						
Ever studied economics	112	0.33 (.)	80	0.45 (.)	-0.12	0.125
Ever participated in experiment	112	0.46 (.)	80	0.71 (.)	-0.25	0.001

Note: Variables are contious when asterisked and binary otherwise. p-values are obtained from a test of equal proportions for binary, and from a t-test for continuous variables. Standard deviations are reported in parantheses for the means of continuous variables. Age and union duration are measured in years, satisfaction with the union is self-reported on a 11-point scale, where 0 indicates not satisfied at all and 10 very satisfied.

5.4 Main results

5.4.1 Do gender norms improve coordination rates?

We begin by examining coordination rates. Figure 5-2 shows the proportion of couples that coordinate, broken down by treatment, for all pairings in the sample (panel 1) and separately for real partners (panel 2) and strangers (panel 3). Coordination rates exceed the theoretical prediction of 45% (shown by the dotted line) in all cases, and, surprisingly, the presence of a focal point improves coordination rates only marginally. On average, the probability of coordinating is only 8 percentage points lower in the *Neutral* treatment. The difference is not statistically significant, as documented by the linear regression results supplied in Table 5.3, column 1.[19] Including a dummy for pairings of strangers, and its interaction with *Neutral* (column 2), confirms that coordination rates do not differ by familiarity of the partners, and do not improve for either group when the labeling of the strategies provides a focal point.

Columns (3), (4), and (5) provide robustness checks. Since coordination occurs at the couple rather than the individual level, the most reasonable approach to analysing this outcome uses couples as observation units. The downside, however, is that we cannot take advantage of the panel structure of the data, since, by design, new couples are formed when subjects are matched to a different partner. Thus, although heteroskedasticity-robust (Huber-White) standard errors can be obtained for these estimations (and are reported in parentheses), they cannot be clustered at the couple level. In columns (3), (4), and (5), I therefore use individual-level data and estimate a pooled OLS model (3) and random effects model (4), with clustered standard errors at the individual level. I use men only to ensure that

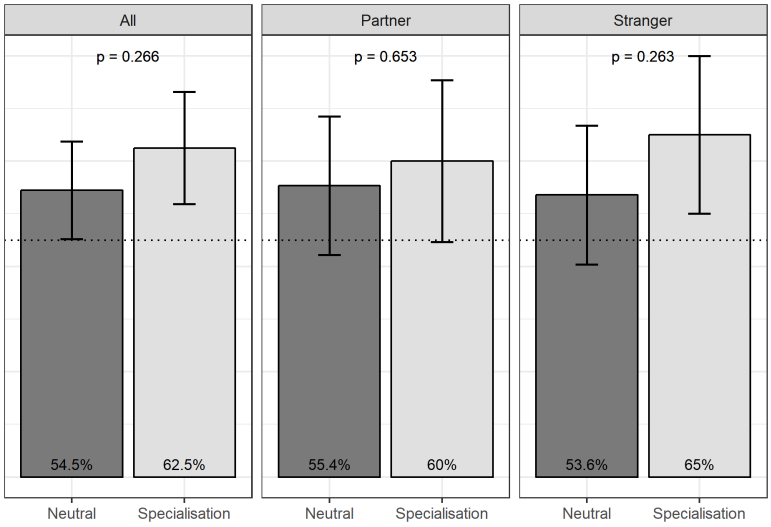
[19] All outcome variables analysed in this chapter are binary and most estimations presented include interaction effects, for which neither the sign nor the statistical significance can be readily interpreted from nonlinear models (see Ai and Norton, 2003). I follow Angrist and Pischke (2008), who argue that OLS is appropriate to measure average treatment effects on limited dependent variables (p.94 ff.).

Table 5.3: Coordination rates

	Outcome: Coordinated				
	(1)	(2)	(3)	(4)	(5)
Neutral	−0.080	−0.114	−0.114	−0.114	−0.114
	(0.057)	(0.082)	(0.103)	(0.103)	(0.110)
Partner		−0.050	−0.050	−0.050	−0.050
		(0.075)	(0.111)	(0.109)	(0.109)
Neutral × *Partner*		0.068	0.068	0.068	0.068
		(0.114)	(0.146)	(0.143)	(0.143)
Intercept	0.625***	0.650***	0.650***	0.650***	0.389
	(0.037)	(0.053)	(0.079)	(0.079)	(0.873)
Model	OLS	OLS	OLS	RE	RE
Observational unit	couples	couples	men	men	men
Standard errors	robust	robust	clustered	clustered	clustered
Controls	No	No	No	No	Yes
Observations	192	192	192	192	192
R^2	0.006	0.008	0.008	0.008	0.045

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01. Standard errors are either Huber-White robust or clustered at the couple level. The full set of controls includes: *Age*, Age^2, *Unionduration*, $Unionduration^2$, *Married*, *Riskaversion*, *Everstudiedecon*, *Everinlab*.

Figure 5-2: Coordination rates across treatments



Note: Proportions, p-values, and 95% confidence intervals (shown by whiskers) are calculated from the linear models 1 (All) and 2 (Partner, Stranger) in Table 5.3. The dotted line shows the theoretical prediction discussed in Section 5.2.

the sample is not artificially inflated. As the outcome occurs at the couple level, the results are nearly identical when using the female sample. Column (5) also includes individual controls: age and age squared, duration of the union and its square, as well as indicators for (i) being married to one's partner, (ii) having ever studied economics, and (iii) having ever participated in a lab experiment. Finally, I also include a survey measure for risk aversion. Due to the length of the experiment, it was not feasible to elicit risk aversion experimentally. Instead, I rely on a questionnaire item validated by Dohmen et al. (2011).

5.4.2 Do gender norms affect coordination outcomes?

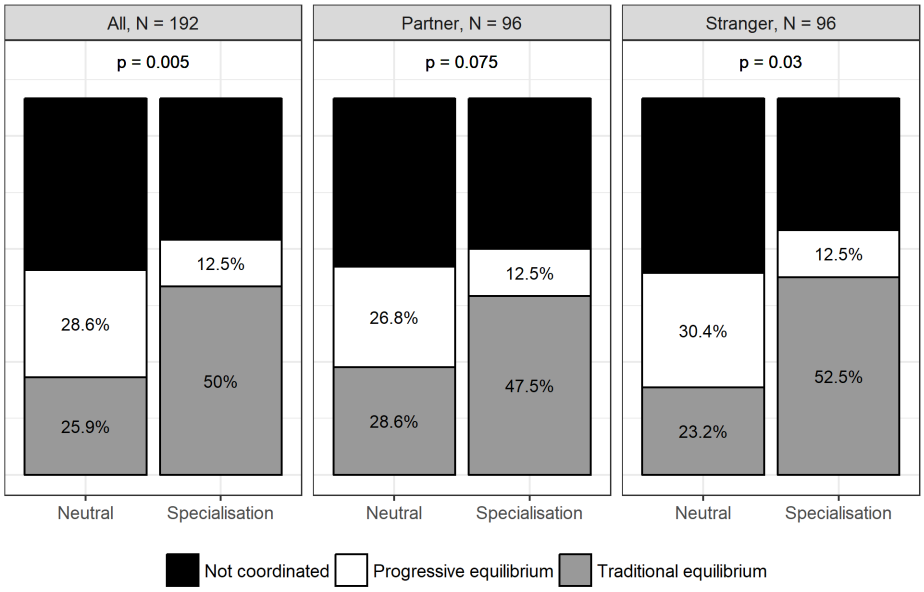
Next, we zoom in and investigate the rate at which each equilibrium is reached across treatments. Despite the small and statistically insignificant differences in coordination rates, Figure 5-3 documents a large difference between the *Neutral* and *Specialisation* treatment in the proportion of couples who reached a *progres-*

sive (i.e., female-favouring) rather than traditional equilibrium. While the two equilibria are nearly equally likely in the *Neutral* control, the traditional equilibrium is reached at a much higher rate in the *Specialisation* treatment. Table 5.4 supplies regression results obtained from linear probability models, which confirm that, on average, the likelihood of reaching a progressive equilibrium is around 16 percentage points higher in the *Neutral* control compared to the *Specialisation* treatment, and is unaffected by familiarity of partners. Because payoffs were higher for the male (female) partner in the traditional (progressive) equilibrium, the male-female earnings gap is larger in the *Specialisation* treatment, to the benefit of male players. On average, women earn only 67% of men's income in the treatment, while in the *Neutral* control both genders earn around 70% of the average male income in the *Specialisation* treatment. Surprisingly, this holds true not just in real couples, in which players may anticipate ex-post redistribution of payoffs, but also for strangers, who cannot redistribute due to anonymity.

5.4.3 Do gender norms affect individual *Career* choices?

Naturally, the differences in the distribution of equilibrium outcomes across treatments shown in result 2 must be brought about by changes in subjects' choices. A priori, one might expect that both men and women adapt their behaviour in the presence of a gender norm focal point, albeit in opposite directions (i.e., compared to the *Neutral* control group, women choose *Career* less often, while men choose it more frequently). If this were the case, however, coordination rates would likely improve, but we have seen in result 1 that this does not happen. To shed light on this puzzle, Figure 5-4 displays the fraction of subjects who choose *Career* across treatments, by gender and partner. As can be seen from the left panel, both genders opt for *Career* at about the same rate in the *Neutral* control, where the overall average is slightly above 50%. Differences across genders and pairings are small and insignificant. The picture changes drastically in the *Specialisation* treatment, where we see a large gender gap in the proportion of subjects who

Figure 5-3: Outcomes across treatments



Note: Proportions, p-values, and 95% confidence intervals (shown by whiskers) are calculated from the linear models 1 (All) and 2 (Partner, Stranger) in Table 5.4, and thus refer to the proportion of couples reaching a progressive equilibrium.

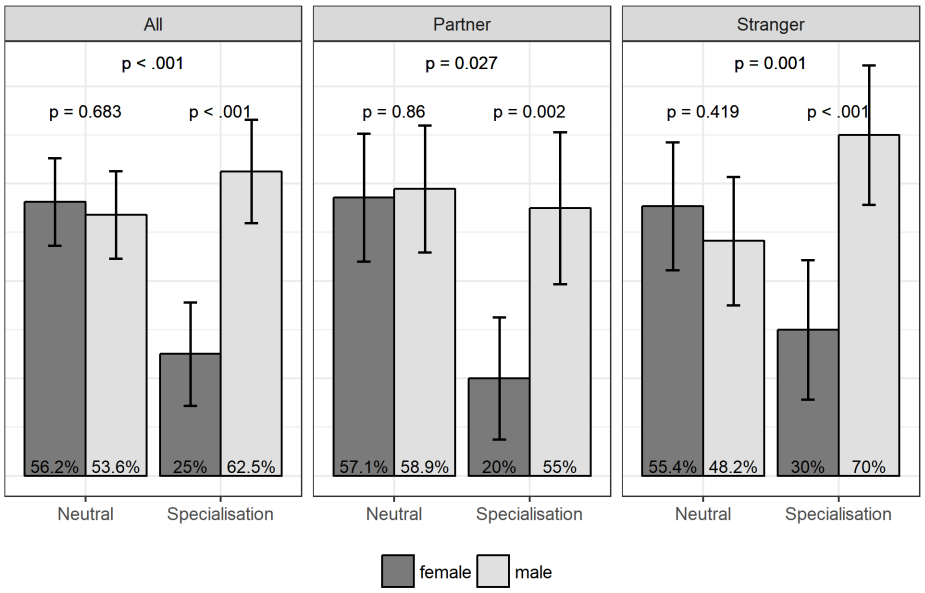
Table 5.4: Progressive equilibrium rates

	Outcome: Progressive equilibrium				
	(1)	(2)	(3)	(4)	(5)
Neutral	0.161***	0.179**	0.179*	0.179*	0.165
	(0.057)	(0.082)	(0.103)	(0.103)	(0.110)
Partner		0.000	0.000	0.000	0.000
		(0.075)	(0.111)	(0.109)	(0.109)
Neutral × *Partner*		−0.036	−0.036	−0.036	−0.036
		(0.114)	(0.146)	(0.143)	(0.143)
Intercept	0.125***	0.125**	0.125	0.125	0.152
	(0.037)	(0.053)	(0.079)	(0.079)	(0.873)
Model	OLS	OLS	OLS	RE	RE
Observational unit	couples	couples	men	men	men
Standard errors	robust	robust	clustered	clustered	clustered
Controls	No	No	No	No	Yes
Observations	192	192	192	192	192
R^2	0.037	0.038	0.038	0.036	0.096

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01. Standard errors are either Huber-White robust or clustered at the couple level. The full set of controls includes: *Age*, Age^2, *Unionduration*, $Unionduration^2$, *Married*, *Riskaversion*, *Everstudiedecon*, *Everinlab*.

choose *Career*. On aggregate, the effect seems to be mainly driven by women, who choose *Career* at a much lower rate than in the control group. The disaggregation by familiarity with the partner, displayed in panels 2 and 3, shows that this happens both when women are paired with their real partner and with a stranger. Moreover, it reveals a remarkable difference in men's choices: Interestingly, men opt for *Career* more frequently compared to the *Neutral* control group only when paired with a stranger, but not when playing with their real partner. Notably, this is not an effect of men generally behaving more 'hawkish' towards women who are not their partners, since men's probability to opt for *Career* in the *Neutral* control does not differ by familiarity with their partner.

Figure 5-4: Proportion of subjects choosing *Career*



Note: Proportions, p-values, and 95% confidence intervals (shown by whiskers) are calculated from the linear models 1 (All) and 2 (Partner, Stranger) in Table 5.5.

The regression results provided in Table 5.5 allow conclusions regarding the statistical significance of the differences in group means displayed in Figure 5-4. Column (1) includes indicators for the *Neutral* control group, female subjects,

Table 5.5: Proportion of *Career* choices

	Choice: *Career*			
	(1)	(2)	(3)	(4)
Neutral	−0.089	−0.218**	−0.218**	−0.194*
	(0.071)	(0.100)	(0.100)	(0.104)
Female	−0.375***	−0.400***	−0.400***	−0.398***
	(0.077)	(0.108)	(0.108)	(0.110)
Partner		−0.150	−0.150	−0.150
		(0.108)	(0.108)	(0.109)
Neutral × *Female*	0.402***	0.471***	0.471***	0.467***
	(0.100)	(0.142)	(0.142)	(0.144)
Neutral × *Partner*		0.257*	0.257*	0.257*
		(0.142)	(0.142)	(0.143)
Female × *Partner*		0.050	0.050	0.050
		(0.153)	(0.153)	(0.154)
Neutral × *Female* × *Partner*		−0.139	−0.139	−0.139
		(0.201)	(0.201)	(0.202)
Intercept	0.625***	0.700***	0.700***	0.724
	(0.054)	(0.077)	(0.077)	(0.514)
Model	OLS	OLS	RE	RE
Controls	No	No	No	Yes
Observations	384	384	384	384
R^2	0.071	0.081	0.081	0.086

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the couple level. The full set of controls includes: *Age*, Age^2, *Unionduration*, $Unionduration^2$, *Married*, *Riskaversion*, *Everstudiedecon*, *Everinlab*.

and their interactions, and thus produces the results displayed in panel 1 of Figure 5-4. The constant reports the proportion of male subjects choosing *Career* in the *Specialisation* treatment (62.5%). In the *Neutral* control group, the share reduces slightly (by 9 percentage points), but the difference is not statistically significant. Among women in the *Specialisation* treatment group, the proportion of those who choose *Career* is significantly smaller than that of men (38 percentage points, i.e., only 25% of women). Women in the *Neutral* control, however, are significantly more likely to choose *Career* than women in the *Specialisation* treatment (40 percentage points). Since the linear combination of the two coefficients *Female* and *Neutral* × *Female* does not differ from zero statistically, men and women in the *Neutral* control group opt for *Career* at the same rate.

Column (2) of Table 5.5 generates the group means disaggregated by familiarity with the partner, which are displayed in panels 2 and 3 of Figure 5-4. Here, the constant reflects the proportion of men in the *Specialisation* treatment opting for *Career* when playing with a stranger (70 %). In the *Neutral* control, the baseline decreases by around 20 percentage points. When playing with their real partner in the *Specialisation* treatment, men's propensity to choose *Career* decreases by 15 percentage points, but the effect is not statistically significant. However, the interaction effect *Neutral* × *Partner* is, which means that the difference-in-difference between choices with the real partner and the stranger *across* treatments is statistically significant. In other words, men in the *Specialisation* treatment choose *Career* at a higher rate when playing with a stranger, but not their partner. For women, we see the same effects as in column 1: They are less likely than men to play *Career* in the *Specialisation* treatment, but not in the *Neutral* control group. Although in Figure 5-4 it appears as if women reduced their propensity to choose *Career* more strongly when playing with their partners as compared to strangers, the difference (obtained by evaluating the linear combination of the interaction *Neutral* × *Partner* and the triple interaction *Neutral* × *Female* × *Partner*) is not statistically significant.

5.5 Mechanisms

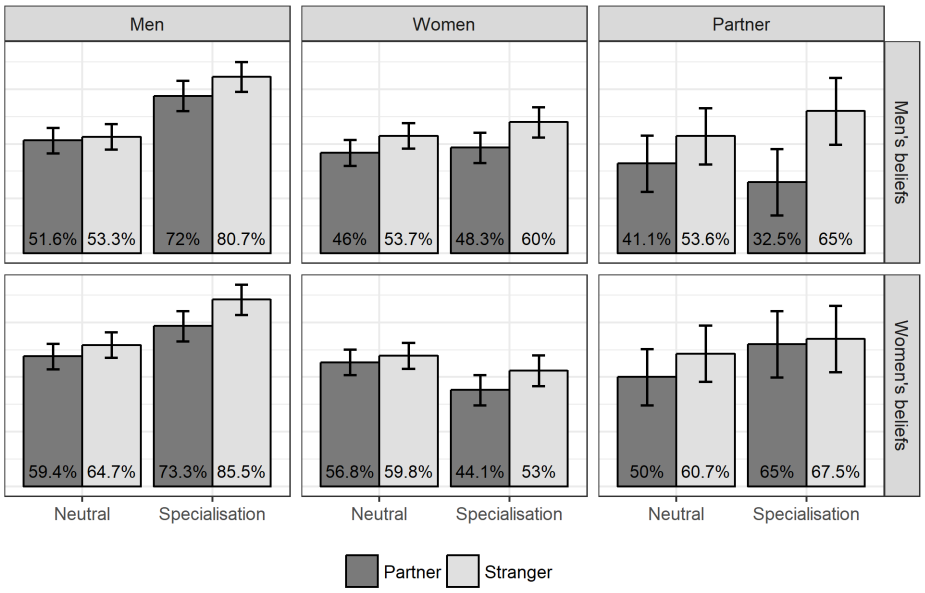
The fact that men that in the *Specialisation* treatment opt for *Career* more frequently only when matched with a stranger, but not when playing with their partner, warrants further investigation. An obvious explanation is differential beliefs regarding their partners' choice. They may expect their real partners to have more progressive norms than the average female student, and thus consider them more likely to choose *Career*. Another (complementary or alternative) reason is a marriage market motive in the spirit of Bursztyn et al. (2017): Men may want to signal progressive gender norms to their partner, but do not care about the impression their choice makes on an anonymous stranger. This section explores these two hypothetical mechanisms empirically.

5.5.1 Differences in beliefs regarding real partner's and stranger's choices?

If beliefs were to explain men's and women's *Career* choices across treatments, women's beliefs about men's behaviour (real partners and strangers) ought to show higher rates of *Career* choices in the *Specialisation* treatment than in the *Neutral* control. Men's beliefs regarding women's choices, however, should not differ between *Specialisation* and *Neutral* with regard to their real partner, but should show lower levels of *Career* choices for the randomly matched women in the *Specialisation* treatment. As detailed in Section 3, participants were asked to predict their partner's choice and the number of women (men) who chose option *Career*/*A* when playing with their partner and when playing with a stranger in their session. For the analyses presented in this subsection, I use these predictions to calculate (i) the proportion of subjects who predicted their partner chose

Career, and (ii) subjects' predicted proportion of men (women) in their session who chose *Career*.[20]

Figure 5-5: Beliefs about *Career* choices across treatments



Note: Proportions, p-values, and 95% confidence intervals (shown by whiskers) are calculated (by column) from the linear models presented in Appendix-Table 5.7.

Figure 5-5 summarises the results; the corresponding regression results are shown in Appendix-Table 5.7. Male subjects' beliefs are displayed in the first row, with females' beliefs in the second row. Column 1 shows the average beliefs about the proportion of men who chose *Career*. The general pattern in men's and women's beliefs looks very similiar; both expect a higher proportion of men to have chosen *Career* in the *Specialisation* treatment compared to *Neutral*, and the difference is statistically significant. ($p < .01$ for all subgroups).[21] While

[20]Predicted proportions are obtained by dividing the absolute number of men (women) that the subject predicted to have chosen *Career* by the total number of male (female) players in her session.

[21]All results obtained from testing linear combinations of the coefficients from the regressions presented in Appendix-Table5.7.

the increase appears to be stronger in pairings of strangers for both genders, the difference in difference is not statistically significant.

Column 2 of Figure 5-5 shows men's and women's average beliefs regarding the proportion of women who choose *Career*. Interestingly, men's beliefs about women's choices are not affected by the treatment, i.e., they do not predict significant differences in women's behaviour, neither for their real partners nor strangers. Male participants do, however, appear to believe women generally behave more 'hawkish' when playing with a stranger as opposed to their real partner; the difference is significant in both treatments ($p < .01$ in both). Women, in contrast, do predict a slightly lower share of women who choose *Career* in the *Specialisation* treatment, although the difference only reaches statistical significance when comparing choices with real partners ($p < .01$). Even so, they too are far from correctly predicting the true rate at which female subjects chose *Career*, which is around 25% in the *Specialisation* treatment (see result 3, Section 5.4). It thus seems that both women and men overestimate (other) women's progressiveness.

Finally, column 3 of Figure 5-5 shows subjects' beliefs regarding their partner's choices, i.e., the proportion of men (women) who believe that their female (male) partner's choice was *Career*. Since the interpretation of these figures differs slightly from those in columns 1 and 2, it is particularly interesting to compare columns 3 and 2 for men (i.e., men's beliefs about the behaviour of their female partners vs. female subjects in their session, respectively) and, analogously, columns 3 and 1 for women. The plot reconfirms that men do not anticipate significant differences in their partners' behaviour across treatments. Again, it appears that men more often think of their stranger partner as 'hawkish' as opposed to their real partner. The difference is particularly pronounced in the *Specialisation* treatment, and is also only statistically significant in this group ($p < .01$). Here, the share of men who *believe* their real partner chose *Career* (33%) actually comes closer to the true share of women who did, while they gravely overestimate it in their stranger partners (65%). Women's beliefs regarding their partners' behav-

ior, on the other hand, seems largely unaffected by familiarity or gender norms, as none of the differences were statistically significant. This is in contrast to the differences in women's beliefs about male subjects' behaviour in their session, as discussed for the panel in column 1. The results in column 3, however, must be taken with a grain of salt, as predictions regarding partner behaviour, unlike the predictions about the general behaviour of men and women, may be subject to hedging motives. It is possible that (some) subjects state the opposite of their true belief to collateralise against coordination failure, provided that they have chosen their strategy in line with their true belief.[22]

Overall, beliefs about choices indicate that subjects do not anticipate gender norms to have a dramatic effect on women's behaviour. Notably, both men and women seem to expect that, in general, mostly men adapt their behaviour in the *Specialisation* treatment and opt for *Career* more frequently than in the *Neutral* control. Thus, while women's choices are broadly consistent with their beliefs, men's beliefs do not seem to resolve the puzzle in their behaviour, i.e., choosing *Career* at a higher rate in the *Specialisation* treatment when playing with a stranger but not with their real partner. The results presented here provide no evidence that this happens because men believe that their real partners are likely to behave more progressively than a randomly drawn stranger from the general population of female partners. On the contrary, it seems that men choose *Career* at a higher rate *despite* their beliefs that female strangers do the same.

5.5.2 Men 'acting feminist'?

This subsection explores the possibility that men's behaviour can be understood as an attempt to signal progressive gender norms, a quality that might be desirable

[22]Note that this concern does not extend to men's predictions regarding the number of women who chose *Career* in their session. Although the difference in predictions regarding women's behaviour toward their real partners and strangers is smaller, we see the same pattern: Men predict women to be less likely to choose *Career* when playing with their real partners as opposed to playing with strangers.

in the marriage market. In a recent paper, Bursztyn et al. (2017) show that single women, unlike women in a domestic relationship or men, avoid certain career-enhancing activities. The authors argue that such activities, while rewarded in the labour market, may be penalised in the marriage market because they signal professional ambition, a quality that men tend to appreciate in their partners only insofar as it remains sufficiently distant from their own Fisman et al. (2006). In conducting several field experiments with Harvard Business School MBA students, Bursztyn et al. (2017) find support for this idea. Single women are 'acting wife' only if their actions are observable to others, and unmarried male peers in particular. For example, they portray themselves as less professionally ambitious (e.g., report significantly lower desired salaries, travel days, and working hours) in an internship placement questionnaire if they expect their answers will be publicly observable compared to those who expect their responses to remain private. The fact that women in a domestic relationship and men do not respond to observability supports the signaling explanation: Unlike single women, neither group stands to gain from marriage market signaling—married women because they already found a match, and men because professional ambition for them is rewarded in both the labour and marriage markets.

In the context of the *battle of the sexes* game analysed in this chapter, it might be men who stand to gain from marriage market signaling by "acting feminist." While choosing *Family* instead of *Career* may reduce the probability of coordinating with their partner, and thus expected earnings, it might signal a desirable quality in the marriage market: progressive gender norms. A recent sociological study (a survey experiment) finds that the majority of college-educated women in the US prefer a progressive over a traditional labour division (Pedulla and Thébaud, 2015). For Germany, Lück (2015) shows, using a 2012 survey on family types, that college-educated respondents are more strongly oriented toward an "active father" rather than a "male breadwinner" model, while the reverse is true

among those without college education.[23] To the extent that men are aware of these (stated) female preferences, they may have an incentive to improve their standing by 'acting feminist' toward a potential long-term partner, i.e., when their relationship is not yet committed. By exploiting a similar setup as Bursztyn et al. (2017), I investigate the effect of observability by their potential long-term partner on men's *Career* choices. If signaling is indeed a motive, men who are in committed relationships should not act differently when their behaviour is observed by their real partner versus a stranger. Men in non-committed relationships, however, may behave differently towards their real partner than towards a stranger, in order to signal progressive gender norms to the former.

I follow Bursztyn et al. (2017) in defining a relationship as serious or committed when partners cohabitate. This allows investigating whether men in non-comimitted relationships, who still compete for their partner in the marriage market, have a desire to signal 'progressiveness' to them, but do not care how their actions are perceived by a random female stranger. I address this question using information on participants' cohabitation status obtained through the post-experimental questionnaire. To facilitate interpretation, and to match the setup in Bursztyn et al. (2017) as closely as possible, I estimate separate regressions for male and female subjects in the different treatments using the following equation:

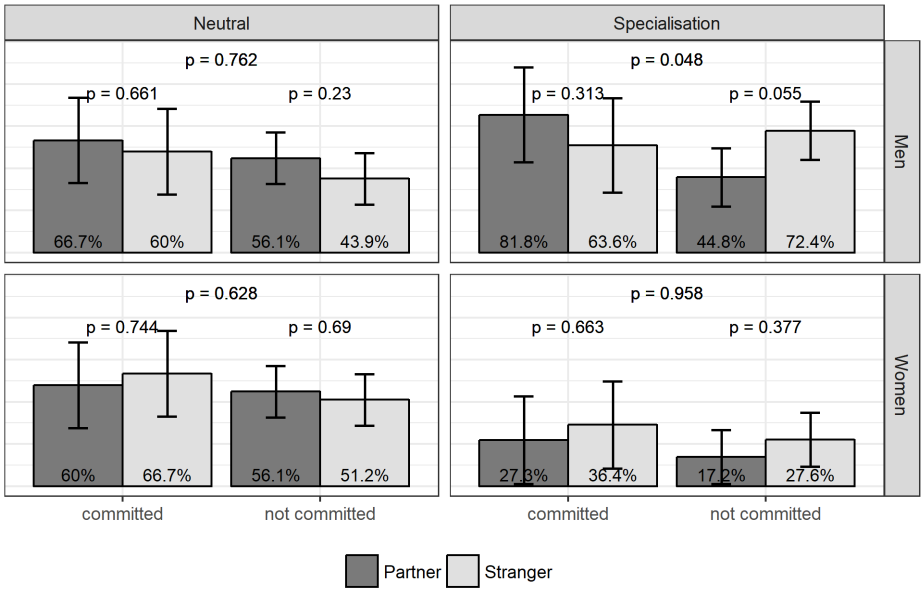
$$\begin{aligned} Career_i = &\beta_0 + \beta_1 Partner + \beta_2 Committed + \\ &\beta_3 Partner \times Committed + \beta_4 \boldsymbol{X_i} + \epsilon_i \end{aligned} \tag{5.1}$$

The vector $\mathbf{X_i}$ holds the usual set of additional controls. Including these controls does not alter the main conclusion; the results are presented in Appendix-Table 5.8. Below, I report the raw coefficients excluding controls. Given this specification, the coefficient β_1 gives the difference in the likelihood of choosing *Career*

[23] This also suggests that coordination rates might actually increase more markedly in the presence of gender norms, if the sample were more representative of the general population in terms of educational background.

when a subject in a non-committed relationship plays with their real partner versus a stranger. Similarly, the coefficient β_2 on the dummy *Committed* gives us the average difference between committed and non-committed subjects playing with strangers. Finally, the coefficient β_3 on the interaction term $Partner \times Committed$ shows the difference in difference in the behaviour change of the two types (committed vs. non-committed) when paired with their partner.

Figure 5-6: *Career* choices by subjects in committed and uncommitted relationships across treatments



Note: Proportions, p-values, and 95% confidence intervals (shown by whiskers) are calculated from the linear models in Table 5.6.

The results are reported in Table 5.6 and support the hypothesis that not choosing *Career*, i.e., choosing *Family* instead, might have signaling value to men who are not in committed relationships because they are less likely to choose *Career* when playing with their real partners and *more* likely when playing with strangers than men in committed relationships. The results are visualised in Figure 5-6. Similar to Bursztyn et al. (2017), and consistent with a signaling interpretation, observability by a potential long-term partner only matters for the behaviour of

Table 5.6: The impact of partner observability on *Career* choices by gender and treatment

	Choice: *Career*			
	Men		Women	
	(1)	(2)	(3)	(4)
Partner	−0.276** (0.125)	0.122 (0.111)	−0.103 (0.115)	0.049 (0.111)
Committed	−0.088 (0.168)	0.161 (0.151)	0.088 (0.156)	0.154 (0.152)
Partner × *Committed*	0.458* (0.238)	−0.055 (0.214)	0.013 (0.220)	−0.115 (0.214)
Intercept	0.724*** (0.088)	0.439*** (0.078)	0.276*** (0.082)	0.512*** (0.078)
Model	OLS	OLS	OLS	OLS
Treatment	*Specialisation*	*Neutral*	*Specialisation*	*Neutral*
Controls	No	No	No	No
Observations	80	112	80	112
R^2	0.085	0.026	0.023	0.010

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the couple level.

non-committed men in the *Specialisation* treatment (top right panel), but does not seem to matter for committed men or women of both types in both treatments.

The signalling interpretation seems to be confirmed when using alternative indicators. With regard to self-reported gender norms, men who hold progressive norms show a similar signalling behaviour in the *Specialisation* treatment. Furthermore, men whose mother was a housewife when they were 15 years old also appear keen on signalling a more progressive view on gender roles to their partners. It is unclear how to exploit relationship duration as an alternative measure for commitment because the costs of separation increase drastically when entering cohabitation. Couples who have been together for a long time, but are not cohabitting might actually be particularly unwilling to commit.

5.6 Conclusion

Economic analyses have recently focused on the potential downsides of gender norms for economic efficiency (Bertrand et al., 2015; Cudeville and Recoules, 2015; Görges, 2015). Households in which women, not men, hold a comparative advantage in labour market production might face a trade-off between maximising monetary income and conforming to gender norms that prescribe, e.g., that 'a man should not do women's work' (Akerlof and Kranton, 2000). This chapter investigates an alternative view of gender norms: their capacity to increase economic efficiency by providing a coordination device for equally productive, non-cooperative spouses.

To this end, 192 subjects in real heterosexual couples, played a classic *battle of the sexes*, i.e., a symmetric coordination game where the two pure strategy Nash equilibria result in unequal payoffs that favour either the male or female partner. The strategies were framed neutrally in the control group (option A vs. B) and as a family specialisation decision in the treatment group (option $Career$

vs. *Family*). Subjects played the game once with their partner and once with a randomly matched stranger.

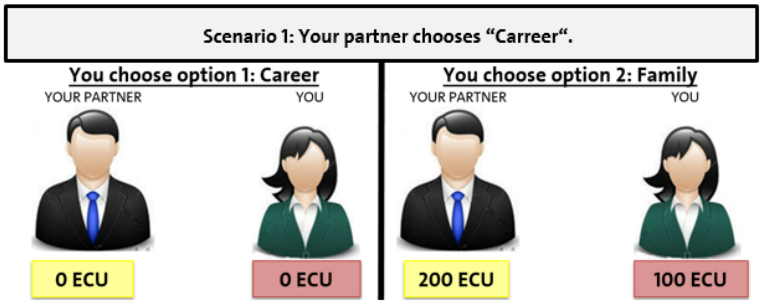
The results provide important insights: First, compared to the *Neutral* control group, overall coordination rates improve only by a small and statistically insignificant margin in the *Specialisation* treatment. This is true for randomly matched strangers as much as for real couples, even though the latter group may revert to post-experimental redistribution of earnings. Thus, there is no evidence that the *presence* of gender norms improves economic efficiency. Second, despite the lack of improvement in overall coordination rates, coordination outcomes change drastically between the *Neutral* and *Specialisation* groups: the traditional gender role equilibrium is reached at a significantly higher rate in the *Specialisation* treatment, i.e., payoff inequality increases to the benefit of male players. Finally, an investigation of individual choices by gender reveals an unexpected pattern: compared to the *Neutral* control group, women opt for *Career* at a significantly lower rate in the *Specialisation* treatment, regardless of familiarity with their partner. Men, however, are only more likely to opt for *Career* when they play with a stranger, but not with their real partner. My results show that men's behaviour does not seem to be explained by beliefs about their partners' choices. Rather, it appears that some men, at the expense of an increased risk of coordination failure, wish to signal progressive gender norms to their partner, a quality that may be desirable in the marriage market.

These results cast doubt on the notion that gender norms serve as an 'efficiency booster,' but support the view that they affect labour market choices of both men and women. The suprisingly heterogeneous effect on men's and women's behaviour, however, suggests that the underlying mechanisms might be more complex than previously considered. This insight calls for further investigation, as a sound understanding of how precisely gender norms affect household behaviour is crucial for designing effective policy strategies that target gender gaps in economic outcomes.

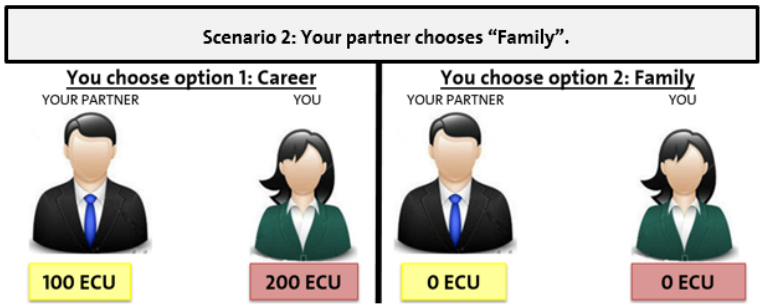
5.7 Appendix A: Experimental instructions

Figure 5-7: Experimental instructions

In this part of the experiment you will choose one of two options: Option 1 („career") or option 2 („family"). Your partner makes the same decision. At the time when you are making your choice, you do not know the choice of your partner. Similarly, your partner does not have any information on your choice either. Your payout , however, depends on the choices both of you make. The result of your decisions will be payed out at the end of this part, i.e. you will find out about each other's choice at that point.



- If both of you choose „Career", each person receives 0 ECU.
- If your partner chooses „Career" and you choose „Family", you receive 100 ECU and your partner 200 ECU.

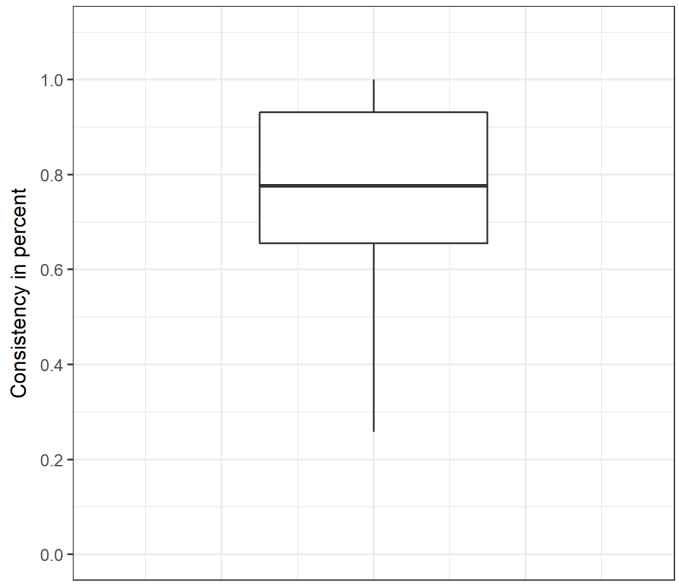


- If your partner chooses „Family" and you choose „Career", you receive 200 ECU and your partner 100 ECU.
- If both of you choose „Family", each person receives 0 ECU.

5.8 Appendix B: Couple consistency check

To alleviate concerns about "fake couples" contaminating the analyses presented in this study, I use the exact same procedure as presented in Chapter 4, Appendix 4.7. The boxplot below shows that the vast majority of couples achieve high consistency rates (obtained by normalising the consistency score). In fact, three quarters of the couples achieve at least 65.5% of the maximum score. The median couple achieves 77.6%. Only 8 couples (fewer than 9%), achieve less than 50% of the score. All analyses presented in the chapter are unaffected if these couples are excluded.

Figure 5-8: Couple's normalised consistency score



Note: The boxplot shows the distribution of couples over the normalised consistncy score. It is calculated by dividing the number of points a couple scored in answering the questions documented in Table 4.6 over the maximum number of points.

5.9 Appendix C: Additional results

Table 5.7: Beliefs regarding men's, women's and partner's *Career* choices, generating Figure 5-5

	Beliefs regarding *Career* choice of ...		
	Men	Women	Partner
	(1)	(2)	(3)
Neutral	-0.274^{***}	-0.062	-0.114
	(0.046)	(0.047)	(0.102)
Female	0.048	-0.070	0.025
	(0.050)	(0.050)	(0.110)
Partner	-0.087^{*}	-0.116^{**}	-0.325^{***}
	(0.050)	(0.050)	(0.110)
Neutral × *Female*	0.066	0.130^{**}	0.046
	(0.065)	(0.066)	(0.144)
Neutral × *Partner*	0.070	0.039	0.200
	(0.065)	(0.066)	(0.144)
Female × *Partner*	-0.035	0.027	0.300^{*}
	(0.071)	(0.071)	(0.155)
Neutral × *Female* × *Partner*	-0.0001	0.020	-0.282
	(0.092)	(0.093)	(0.203)
Intercept	0.807^{***}	0.600^{***}	0.650^{***}
	(0.035)	(0.036)	(0.078)
Model	OLS	OLS	OLS
Standard errors	clustered	clustered	clustered
Controls	No	No	No
Observations	384	384	384
Adjusted R^2	0.198	0.042	0.033

Note: Stars indicate significance $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$. Standard errors are clustered at the couple level.

Table 5.8: The impact of partner observability on *Career* choices by gender and treatment, including controls

	Choice: *Career*			
	Men		Women	
	(1)	(2)	(3)	(4)
Partner	−0.276** (0.122)	0.122 (0.110)	−0.103 (0.119)	0.049 (0.114)
Committed	0.023 (0.176)	0.212 (0.189)	0.111 (0.172)	0.130 (0.203)
Partner × *Committed*	0.458** (0.232)	−0.055 (0.213)	0.013 (0.226)	−0.115 (0.220)
Intercept	−2.393* (1.278)	1.326 (1.559)	0.526 (0.899)	1.366 (2.153)
Model	OLS	OLS	OLS	OLS
Treatment	*Specialisation*	*Neutral*	*Specialisation*	*Neutral*
Controls	Yes	Yes	Yes	Yes
Observations	80	112	80	112
Adjusted R^2	0.093	0.009	−0.077	−0.075

Note: Stars indicate significance *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the couple level. The full set of controls includes: *Age*, Age^2, *Unionduration*, $Unionduration^2$, *Married*, *Riskaversion*, *Everstudiedecon*, *Everinlab*.

General Conclusion

This final chapter conflates the principal findings of the thesis and presents a discussion of policy implications. Finally, it also offers an outlook that sketches potential pathways for future research.

Principal findings and policy implications

Gender differences in participation in labour market and home production activities contribute to persisting gender inequality in economic outcomes, an important concern to policymakers. The chapters of this thesis scrutinize the manifold ways in which gender norms perpetuate the gendered patterns in the aforementioned economic activities. While it is widely acknowledged that gender norms do play a role in the unequal participation in market and family work between men and women, among both policymakers and economic researchers, far too little is known about the precise channels through which they impact behaviour to inform public policy strategies. The studies outlined in this thesis provide a step toward addressing this knowledge gap; however, as will be discussed in the next subsection, much work still remains.

Overall, the chapters in this thesis show that public policy has the potential to alter gender norms, thereby allowing for significant improvements toward equality. The analyses in Chapter 1 suggest that policies may shift gender differences in preferences by actively encouraging men and women to take on new roles. Exam-

ples for such policies include gender quotas in management positions and paternity leave benefits, which may not only reduce the gender gap in participation mechanically, but also by changing men's and women's aspirations to these roles.

Chapter 2 suggests that language may impact gender identity, thereby contributing to gender-stereotypical labour market choices of men and women. While it also shows that measuring a causal effect empirically is highly challenging, the theoretical analysis suggests a promising avenue for implementing public policies to reduce the impact of gender norms on behaviour, particularly in countries where the dominant language is highly gendered, such as Germany. Promoting gender-inclusive language by routinely referring to both male and female forms, e.g., businessmen and businesswomen, may increase the mental representation of both genders in jobs and activities that are not typically female- or male-dominated. This rather low-cost intervention may reduce normative prescriptions of what is deemed an appropriate activity for each gender considerably.

Chapters 3–5 invoke a family economic perspective and experimental methods to study the effects of gender norms on labour division. The main conclusions here concern the question of whether norms—apart from fueling gender differences in preferences, aspirations, and productivity—may also impact labour division by imposing psychological, "identity" costs on deviant behaviour. This economic view on gender identity as a good that households may have some demand for immediately points to policy instruments that subsidise the consumption of gender identity, e.g., through differential taxation of spouses, as enabling gender norms to sustain inequality. I find some empirical evidence for the identity hypothesis. For example, in Chapter 4, I show that men supply more effort to the market when the threat of being out-earned by their partner is high. While in the experimental setting this does not necessarily come at a cost, a similar response in the real world would imply that men have fewer time resources to invest in family work, and the shortage is likely to be asorbed by women. This type of behaviour may hinder the

effectiveness of policy instruments aimed to promote the equality of partners, and should thus be taken into account when designing such measures.

The evidence presented in Chapters 3 and 4 also suggests that identity may not be the only motive that, even in the absence of gender differences in productivity and preferences, leads to gendered specialisation. Furthermore, uncertainty about future wages and career developments presents another important factor for sustaining traditional gender specialisation patterns because men are more likely to be overconfident in their earning prospects than women. This finding, too, suggests that the adoption of gender quotas can confer benefits, as they may raise women's perceived probability of having a successful career relative to their partners. Moreover, Chapter 4 identifies yet another channel contributing to the prevalence of traditional gender roles: women's reluctance to assume sole responsibility for family earnings. To the extent that specialisation per se, for its potential to raise the productive efficiency of households, is considered desirable by society, policymakers should find ways to reduce women's reluctance to assume breadwinner responsibilities; otherwise, specialisation patterns will likely remain gendered.

Finally, Chapter 5 highlights yet another difficulty in combating gender norms. Couples may, to some extent, perceive them as useful for resolving conflicts of interest, and thus may resist their attentuation. The study shows that, suprisingly, even young female university students retreat from more lucrative career activities to the family domain and claim a smaller part of the pie for themselves when gender norms suggest this may be the "right" allocation. Men's behaviour is more heterogeneous, and appears to reflect a motive of signalling progressive gender norms to a courted partner. This points to a further challenge for policymakers, as it suggests that partners, especially at the beginning of a relationship, may not have (or show) any intentions to specialise according to traditional gender norms. As a result, encouraging an open and deliberate bargaining process about the allocation of breadwinner and homemaker roles might be further complicated—

but is also indispensable given that those roles are not only afflicted by norms but also induce asymmetric risks for partners.

Outlook

This final subsection briefly sketches potential avenues for future research. Chapter 1 provides evidence that gender differences in preferences are malleable by institutions, as shown by the example of GDR institutions, which actively promoted the role of women as workers. We argue that policies such as gender quotas in management positions and parental leave benefits reserved for fathers can bring about similar shifts, as they encourage men and women to take on new roles. Naturally, this prompts the question of whether these policies, where introduced, led to comparable effects on gender differences in preferences. Of particular interest is whether such effects, if detectable, are confined to individuals directly affected by the policy or if they extend further; e.g., do men with no children become more likely to consider the role of caretaker as suitable for men?

Chapter 2 identifies significant challenges for research that aims to measure a direct, causal impact of gendered language on speaker behaviour, and uses an empirical application to illustrate the problem. The solution is left to future research. One potentially promising avenue is to study the exogenous assignment of languages to speakers when they are very young, as is the case for child migrants who were brought to a different country by their parents. Within Europe, it is possible to exploit the fact that countries differ with respect to gender marking in the dominant language; thus, migrant children receive differential treatment depending on where they arrive. Another possibility would be to study the effect of a highly gendered language (e.g., German) versus a less gender-intensive language (e.g., English) on gender gaps in economic choices within a laboratory setting.

Chapter 3 presents the first study of spousal labour division choices and shows that gender identity may be an important factor. Chapter 4 builds on this finding and sets out to investigate the identity channel more rigorously. The specialisation

game is adapted so that it no longer poses an asymmetric risk to the disadvantage of subjects specialising in homemaking. However, I believe that the *hold-up problem* (Hart, 1995) implicit in the specialisation decision in the Chapter 3 study deserves further attention. The gendered specialisation patterns this study finds in real couples, but not in pairs of strangers, may indicate that women have a greater willingness to accept income asymmetry to their disadvantage when paired with a romantic partner. This hypothesis is, to some extent, also supported by the findings in the battle of the sexes study in Chapter 5, and thus warrants further investigation. Ultimately, the question of whether marriage contracts can remedy the hold-up problem, or are rejected by couples because they violate a norm that "true love does not need contracting," is an important and promising direction for future research.

Another important question that has not been addressed in the experimental studies in this thesis is how the impact of gender norms on intra-household labour division interacts with families' socioeconomic background or other household characteristics. This is especially salient for those chapters that rely on an identity framework, which implicitly or explicitly assume that conformity with gender norms is an immaterial good that the household may or may not demand. As is the case for any other good, price increases—in this case, increases in female spouses' relative labour market productivity—trigger income and substitution effects, and size and direction of the net effect will likely depend on overall household income. Other household characteristics may also foster heterogeneous responses to gender norms, e.g., the duration of the union, and factors determining spouses' intra-household bargaining power. The investigation of such aspects is naturally limited by the homogeneity of student samples with respect to such characteristics. Therefore, I believe that a fruitful approach to family economics research based on experimental methods is to conduct survey experiments within the infrastructure of large, established household panels. This would also shed light on the external

validity of the results obtained from experiments with student couples, which at this point is still unexplored.

To summarise, the five essays presented in this thesis constitute an attempt to study different channels through which gender norms affect gender differences in labour market and family work participation using economic methods. While one of the contributions of the thesis lies in the careful distinction of these channels and the systematic assessment of their relevance, it makes no claim that the set of answers provided is exhaustive. Rather, the thesis points to challenges in, and offers new approaches to, understanding the complex interplay of economic considerations and gender norms in couples' labour division decisions.

Summary

This dissertation comprises five studies on the impact of gender norms on female labour supply. Based on theoretical, experimental, and micro-econometric analyses, it aims to provide insights regarding the various channels through which gender norms may affect individual time allocation to labour market and family work.

An introductory chapter explains the common research agenda and presents the unifying theoretical framework that interconnects the independent studies. The chapters then proceed to study different channels through which norms may affect female labour supply in an economic framework, treating individuals as single decision makers at first (Chapters 1 and 2), and subsequently widening the perspective to account for interdependencies in the family (Chapters 3, 4, and 5). A concluding chapter summarises the principal findings and discusses policy implications, before sketching potential pathways for further research. I briefly summarise each chapter below.

Chapter 1 studies the role of social norms and institutions in forming gender differences in preferences using micro-econometric techniques. The identification relies on the German division and reunification as a natural experiment, thus allowing to test whether the two political systems led to different gender gaps in preferences for work. The analysis relies on German-General-Social-Survey data from 1991, 1998, and 2012, augmented by a comprehensive set of register data, as well as an extensive compilation of historical data from the 19th and early 20th

century. The results show a substantial "gap in the gap" across regions, with a smaller difference in the East. This gap in the gap exists directly after reunification and does not close until 2012. It is driven by cohorts of individuals who grew up during separation, and gained experience with GDR institutions that supported female employment; this seems to have built an important gateway for social norms to shape preferences.

Chapter 2 investigates a specific channel through which social norms may impact the value an individual places on gender identity, thereby affecting labour supply decisions: language. The chapter presents (i) a theoretical model for understanding the cognitive effect of language on labour supply based on insights from (psycho-)linguistics and identity economics; (ii) a systematic review of the existing empirical literature that identifies challenges for establishing causality; and (iii) a rigorous empirical assessment of these challenges using data for European labour markets. A key insight derived from the model regards the methodological task of separating the cognitive effects of language from those of institutions and other cultural influences, which remain largely unresolved in the existing empirical literature. The empirical part of the paper illustrates the severity of the problem by employing the epidemiological approach, as a link between behaviour and speaking a gendered language can hardly be detected empirically, regardless of the specification. Overall, the chapter exposes the challenges that arise in studying the impact of gender norms on behaviour, especially when relying on micro and survey data, and a restricted theoretical framework treating individuals as single decision makers. To address these issues, the subsequent chapters build on family economic theory and use experimental methods.

Chapter 3 examines couples' specialisation decisions and the gender-specific patterns in labour division using a lab experiment. Eighty participants—20 real couples and 20 pairs of strangers—play a specialisation game. Partners decide jointly if and how they specialize: They can both complete a performance-based paid task (task A) or have one of the players perform an unpaid task (task B),

thereby tripling the pay rate for the partner performing task A. After completing their tasks, participants are informed about their pay-outs in private and then asked to make an individual decision about what proportion of their income to pay into a common pool, where it is increased by 20% and distributed equally between the two players. The results show that couples are significantly more likely to choose women as performers of the unpaid task than men. In pairs of strangers, no such difference can be detected; women and men are equally likely to perform the task. The results may be driven by couples' desire to conform to gender identity, a hypothesis tested more rigorously in the next chapter.

Chapter 4 presents a systematic approach to understanding how gender norms affect couples' labour division. I develop a theoretical model of time allocation in the presence of gender identity concerns to identify three different channels: gender differences in preferences and productivity, and identity concerns. The model shows that the latter channel may be difficult to identify with survey data since the former two channels are confounders that cannot be isolated. Building on this insight, the experiment is designed to neutralise gender differences in home production and in preferences for the home-produced good. Because specialisation according to comparative advantage in market production is income-maximizing by design, women and men should be chosen as breadwinners with equal probability if they have the same market productivity. However, examining the choices of 246 participants (real heterosexual couples), I find that men are significantly more likely chosen as breadwinners than women. Unpacking the mechanisms, much of the difference stems from gender differences in market productivity. Women, on average, performed worse than men in the market task prior to specialisation. Additionally, men's overconfidence and women's reluctance to assume sole responsibility for income provision appear to amplify the gender representation gap among breadwinners. While I do not find much evidence for a direct effect of gender identity on the labour division choice, and thus on gender differences in labour supply at the extensive margin, effort provision of men and women sug-

gests identity may affect the intensive margin. Consistent with the behavioural prescription "A man should earn more than his wife," I find that men, but not women, supply significantly more effort when their partner was previously more productive, i.e., when the threat that she might out-earn him is high.

Chapter 5 studies the potential benefits of gender norms as a coordination device and efficiency enhancer in the lab. A total of 192 subjects (96 real heterosexual couples) played a classic battle of the sexes game, i.e., a symmetric coordination game where the two pure-strategy Nash equilibria result in unequal payoffs that favour either the male or female partner. The strategies were framed neutrally in the control group (option A vs. B) and as a family specialisation game in the treatment group (option $Career$ vs. $Family$). Subjects played the game once with their partner and once with a randomly matched stranger. The results are surprising in three respects: First, overall coordination rates improve only by a small margin in the *Specialisation* treatment, while payoff inequality increases dramatically to the benefit of male players. That is, the probability that couples coordinate on the traditional gender role equilibrium increases, yet the overall probability of coordinating, and thus efficiency, does not improve much. Second, the effect is equally present among real couples and randomly matched strangers, where post-experimental redistribution of earnings is not possible. Finally, an investigation of individual choices by gender reveals an unexpected pattern: Compared to the control group, women opt for $Career$ at a significantly lower rate in the *Specialisation* treatment, regardless of familiarity with their partner. Men, however, are only more likely to opt for $Career$ when they play with a stranger, but not with their real partner. I present evidence suggesting that some men's reluctance to choose $Career$ when playing with their partner is driven by their desire to signal progressive gender norms.

German Summary (Zusammenfassung)

Gegenstand der vorliegenden Arbeit ist der Einfluss von Geschlechternormen auf die Erwerbsbeteiligung von Frauen. Sie präsentiert fünf ökonomische Studien die auf theoretischen, experimentalökonomischen und mikroökonometrischen Analysen beruhen. Die Studien verbindet das übergreifende Ziel, unterschiedliche Wirkungskanäle zu beleuchten, durch die Normen die Zeitverwendung von Individuen beeinflussen, insbesondere die Aufteilung von Zeitressourcen auf Erwerbs- und Familienarbeit.

Ein einleitendes Kapitel legt das gemeinsame Forschungsinteresse dar und führt einen modelltheoretischen Rahmen ein, der die fünf eigenständigen Studien verbindet. Diese werden in den fünf folgenden Hauptkapiteln vorgestellt. Die Frage, wie Normen Entscheidungen beeinflussen, wird zunächst aus Perspektive einzelner Akteur*innen beleuchtet (Kapitel 1 und 2) und in den anschließenden Kapiteln mit Blick auf die Verflechtungen von Akteur*innen innerhalb der Familie (Kapitel 3, 4 und 5). Im letzten Kapitel werden die wichtigsten Ergebnisse zusammengefasst und einige Schlussfolgerungen zur Diskussion gestellt. Sie beziehen sich auf politische Handlungsmöglichkeiten und verweisen auf Anknüpfungspunkte für die zukünftige Forschung. Der folgende Überblick bietet eine jeweils kurze Zusammenfassung der Hauptkapitel.

Kapitel 1 präsentiert eine mikroökonometrische Studie zum Einfluss sozialer Normen und Institutionen bei der Ausbildung von geschlechtsspezifischen Präferenzen. Die gewählte Forschungsstrategie nutzt die deutsche Teilung und anschließende Wiedervereinigung als natürliches Experiment. Die Studie leistet einen Beitrag zur genaueren Klärung der Frage, ob und in welcher Weise die gesellschaftlichen und politischen Systeme in den beiden deutschen Staaten BRD und der DDR unterschiedliche Geschlechterdifferenzen in Präferenzen für Erwerbsarbeit beeinflusst haben. Die Analyse erfolgt anhand dreier Querschnitte aus der Allgemeinen Bevölkerungsumfrage (ALLBUS) aus den Jahren 1991, 1998 und 2012, ergänzt um offizielle Registerdaten und eine umfangreiche Zusammenstellung historischer Daten aus dem 19. und frühen 20. Jahrhundert. Die Ergebnisse zeigen eine beachtliche „Lücke in der Lücke“ zwischen den beiden Regionen, da der Geschlechterunterschied in den Präferenzen für Arbeit im Osten deutlich kleiner ausfällt als im Westen. Die regionale Differenz zeigt sich direkt nach der Wiedervereinigung und bleibt bis 2012 erhalten. Die Unterschiede zeigen sich vor allem bei den Geburtenjahrgängen, die während der Teilung eigene Erfahrung mit den Institutionen der DDR bzw. BRD gesammelt haben, die die Erwerbsbeteiligung von Frauen förderten bzw. hemmten. Institutionen erweisen sich daher als wirkungsvolle Instrumente, die Geschlechternormen und Präferenzbildungen beeinflussen.

In Kapitel 2 wird ein weiterer Kanal untersucht, über den Geschlechternormen das individuelle Arbeitsangebotsverhalten beeinflussen können: die Sprache. Die hier vorgestellte Studie präsentiert (i) ein theoretisches Modell, das Erkenntnisse aus der (Psycho-) Linguistik und der Identitätsökonomik zusammenführt, um den Effekt von Sprache auf die Arbeitsangebotsentscheidung zu analysieren; (ii) eine systematische Übersicht über die bisherigen ökonomischen Studien zu diesem Thema und über die bislang ungelösten Probleme für den Nachweis eines Kausalzusammenhanges; (iii) eine umfassende empirische Auswertung mit Blick auf die zuvor herausgearbeiteten Probleme, basierend auf europäischen Daten. Eine wichtige Erkenntnis aus dem theoretischen Modell betrifft die – bislang un-

gelöste – methodische Herausforderung, die kognitiv vermittelte Wirkung von Sprache auf Verhalten, von anderen, vor allem institutionellen und kulturellen Einflüssen empirisch zu trennen. Die empirische Analyse verdeutlicht das Problems unter Anwendung des epidemiologischen Ansatzes; ein Zusammenhang zwischen Sprache und Verhalten ist kaum messbar, unabhängig von der verwendeten Spezifikation. Der Aufsatz resümiert die erheblichen theoretischen und methodischen Herausforderungen für die Untersuchung des Einflusses von Geschlechternormen auf das Verhalten. Diesen Herausforderungen ist unter Verwendung von Mikro- und Umfragedaten, sowie eines reduzierten Analyserahmens, der die Akteur*innen als isolierte Individuen betrachtet, nur schwer beizukommen. Um ihnen angemessen zu begegnen, nutzen die folgenden Kapitel familienökonomische Theorien und Experimente.

Kapitel 3 präsentiert Ergebnisse aus einem Laborexperiment, in dessen Rahmen die Spezialisierungsentscheidungen von Paaren und die geschlechtsspezifischen Muster in der Arbeitsteilung untersucht wurden. Hierfür nahmen 80 Teilnehmerinnen und Teilnehmer – 20 echte und 20 einander zufällig zugeteilte fremde Paare – an einem Spezialisierungsspiel teil. Die Paare entschieden gemeinsam, ob und wie sie sich spezialisieren möchten: Entweder führten beide, Partner und Partnerin eine mit Stücklohn entlohnte Aufgabe aus, oder nur einer von beiden tat dies, während der oder die andere sich einer unbezahlten Aufgabe widmete. Nach Abschluss der Aufgaben wurden die Teilnehmenden getrennt und vertraulich über ihre Auszahlung informiert und erhielten die Möglichkeit, ihre Einnahmen in einen gemeinsamen Fonds zu investieren, in dem alle Einzahlungen um 20% gesteigert und anschließend zu gleichen Teilen an den Partner und die Partnerin ausgeschüttet wurden. Im Ergebnis zeigt sich, dass die echten Paare mit signifikant höherer Wahrscheinlichkeit der Partnerin und nicht dem Partner die unbezahlte Aufgabe zuweisen. Demgegenüber führen bei den einander zufällig zugeteilten Paaren die Männer und Frauen die unbezahlte Aufgabe mit gleich hoher Wahrscheinlichkeit aus. Eine mögliche Erklärung für dieses Ergebnis liefert die ökonomische Identität-

stheorie, nach der echte Paare bei der Entscheidung über die partnerschaftliche Arbeitsteilung bestrebt sind, sich gemäß den gesellschaftlichen Geschlechternormen zu verhalten. Diese Hypothese wird im folgenden Kapitel genauer überprüft.

Kapitel 4 widmet sich demzufolge der Frage, auf welche Weise Geschlechternormen die Arbeitsteilung von Paaren beeinflussen. Auf der Grundlage eines theoretischen Modells zur Zeitverwendung, das den Einfluss der Geschlechter-Identitätsverunsicherung explizit berücksichtigt, werden zunächst unterschiedliche Kanäle bestimmt, über die Normen wirken können. Das Modell zeigt, dass Identitätsverunsicherung als Motiv nur schwer anhand von Survey Daten zu erfassen ist, insbesondere wenn Geschlechterunterschiede in Markt- oder Haushaltsproduktivität oder in Präferenzen für die im Haushalt erstellten Güter (wie Kinderbetreuung) bestehen. Das vor diesem Hintergrund entwickelte Laborexperiment erlaubt es, die theoretisch herausgearbeiteten Wirkungskanäle von Normen voneinander zu abzugrenzen. Das Design stellt sicher, dass die Geschlechterunterschiede bezüglich der Präferenzen und der Haushaltsproduktivität die Entscheidungen über die Arbeitsteilung im Labor nicht beeinflussen. Da Spezialisierung gemäß des Produktivitätsvorteils in der Marktarbeit per Design einkommensmaximierend wirkt, müssten die Teilnehmerinnen und Teilnehmer mit gleicher Wahrscheinlichkeit Frauen und Männer für die Marktarbeit wählen, sofern hierin keine systematischen Produktivitätsunterschiede bestehen. Die Analyse der Entscheidungen von 246 Teilnehmerinnen und Teilnehmern (echten heterosexuellen Paaren) zeigt jedoch, dass Männern signifikant häufiger als Frauen die Rolle des Einkommensverdieners zugewiesen wird. Als Hauptursache der Asymmetrie erweist sich, dass Frauen in den der Spezialisierungsentscheidung vorausgehenden Einzelarbeitsrunden bei der Marktaufgabe häufig weniger produktiv waren als ihre Partner. Weitere wichtige Einflussfaktoren sind die Neigung der Männer zu übermäßigem Selbstvertrauen und die Abneigung der Frauen, die alleinige Verantwortung für das Familieneinkommen zu übernehmen. Während sich in den Arbeitsteilungsentscheidungen selbst – und damit in den Entscheidungen darüber, ob Frauen überhaupt im Arbeitsmarkt ak-

tiv sind – kaum Evidenz für einen *direkten* Einfluss der Geschlechteridentität zeigt, deuten Unterschiede im Arbeitseinsatz auf männliche Identitätsverunsicherung hin: In Übereinstimmung mit der Verhaltenszuschreibung einer Geschlechternorm, die besagt „Ein Mann sollte mehr verdienen als seine Frau“, erhöhen Männer ihren Arbeitseinsatz signifikant, wenn ihre Partnerin in vorhergehenden Runden produktiver war als sie selbst.

Kapitel 5 beleuchtet mithilfe eines Laborexperimentes den potenziellen Nutzen von Geschlechternormen als Koordinationshilfe und Motor zur Realisierung von Effizienzgewinnen. 192 Teilnehmerinnen und Teilnehmer (96 Paare) spielten ein klassisches „Kampf der Geschlechter“-Spiel, d.h. ein symmetrisches Koordinationsspiel, in dem die zwei Gleichgewichte in reinen Strategien entweder den Partner oder die Partnerin im Vergleich zum jeweils anderen besserstellen. In der Kontrollgruppe sind die Strategien neutral benannt (Option A versus B) und in der Behandlungsgruppe als Alternativen in der familiären Spezialisierung (Option $Karriere$ versus $Familie$). Im Experiment spielten die Teilnehmenden sowohl mit ihrem Partner bzw. ihrer Partnerin als auch mit einer zugelosten fremden Person. Die Ergebnisse sind in dreierlei Hinsicht überraschend: Erstens zeigt sich, dass die Koordinationsraten im Spezialisierungs-Treatment kaum über denen in der Kontrollgruppe liegen, obwohl im ersteren die Geschlechterungleichheit bei den Auszahlungen deutlich zugunsten der männlichen Teilnehmer zunimmt. Mit anderen Worten: Obwohl die Wahrscheinlichkeit steigt, das „traditionelle“ Gleichgewicht zu erreichen, erhöht sich nur unwesentlich die Wahrscheinlichkeit, dass es überhaupt zu einer Koordination kommt. Zweitens zeigt sich, dass dieser Befund sowohl für echte Paare als auch für einander zufällig zugeteilte fremde Personen gilt. Drittens zeigt die Analyse der individuellen Entscheidungen ein unerwartetes Muster, das zugleich eine Erklärung nahelegt: Im Vergleich zur Kontrollgruppe wählen die Frauen im Spezialisierungs-Treatment deutlich häufiger die Alternative $Familie$, unabhängig davon, ob sie mit ihrem Partner oder einem fremden Mann spielen. Männer hingegen wählen $Karriere$ im Spezialisierungs-Treatment

nur dann häufiger als in der Kontrollgruppe, wenn sie mit einer fremden Frau spielen, nicht jedoch mit ihrer Partnerin. Eine Analyse der möglichen Motive lässt vermuten, dass die Wahl mancher Männer von dem Wunsch getrieben ist, der eigenen Partnerin fortschrittliche Vorstellungen zum Geschlechterverhältnis zu signalisieren.

List of Tables

List of Figures

Bibliography

Abramitzky, R., L. P. Boustan, and K. Eriksson (2014). A nation of immigrants: Assimilation and economic outcomes in the age of mass migration. *Journal of Political Economy 122*(3), 467–506.

Adams, R., R. Kräussl, M. Navone, and P. Verwijmeren (2017). Is gender in the eye of the beholder? Identifying cultural attitudes with art auction prices. Research papers, SSRN. Available online: `http://dx.doi.org/10.2139/ssrn.3083500` (accessed May 18, 2018).

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*(2), 293–337.

Adler, M. A. and A. Brayfield (1997). Women's work values in unified Germany. Regional differences as remnants of the past. *Work and Occupations 24*(2), 245–266.

Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics Letters 80*(1), 123–129.

Aisenbrey, S., M. Evertsson, and D. Grunow (2009). Is there a career penalty for mothers' time out? A comparison of Germany, Sweden and the United States. *Social Forces 88*(2), 573–605.

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

Akerlof, G. A. and R. E. Kranton (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 1167–1201.

Albanesi, S. and C. Olivetti (2016). Gender roles and medical progress. *Journal of Political Economy 124*(3), 650–695.

Alesina, A. and N. Fuchs-Schündeln (2007). Good-bye Lenin (or not?): The effect of communism on people's preferences. *American Economic Review 97*(4), 1507–1521.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alesina, A., A. Ichino, and L. Karabarbounis (2011). Gender-based taxation and the division of family chores. *American Economic Journal: Economic Policy 3*(2), 1–40.

Amato, P. R. and A. Booth (1995). Changes in Gender Role Attitudes and Perceived Marital Quality. *American Sociological Review 60*(1), 58–66.

Ambler, K. (2015). Don't tell on me: Experimental evidence of asymmetric information in transnational households. *Journal of Development Economics 113*, 52–69.

Anderberg, D., H. Rainer, J. Wadsworth, and T. Wilson (2016). Unemployment and domestic violence: theory and evidence. *The Economic Journal 126*(597), 1947–1979.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Apps, P. and R. Rees (2009). *Public economics and the household*. Cambridge University Press.

Apps, P. F. and R. Rees (1988). Taxation and the household. *Journal of Public Economics 35*(3), 355–369.

Arrow, K. (1972). Models of job discrimination. In A. Pascal (Ed.), *Racial discrimination in economic life*, pp. 83–102. Lexington Books.

Asher, R. E. and C. Moseley (1994). *Atlas of the world's languages*. Routledge.

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

Ashraf, N., E. Field, and J. Lee (2014). Household bargaining and excess fertility: An experimental study in Zambia. *American Economic Review 104*(7), 2210–2237.

Attanasio, O. and V. Lechene (2002). Tests of income pooling in household decisions. *Review of Economic Dynamics 5*(4), 720–748.

Baker, M. J. and J. P. Jacobsen (2007). Marriage, specialization, and the gender division of labor. *Journal of Labor Economics 25*(4), 763–793.

Basu, K. (2006). Gender and say: A model of household behaviour with endogenously determined balance of power. *The Economic Journal 116*(511), 558–580.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48.

Bauernschuster, S. and H. Rainer (2012). Political regimes and the family: How sex-role attitudes continue to differ in reunified Germany. *Journal of Population Economics 25*(1), 5–27.

Beblo, M. (2001). *Bargaining over time allocation: Economic modeling and econometric investigation of time use within families*. Physica Verlag.

Beblo, M. and D. Beninger (2010). Dokumentation zur Studie "Konsumentscheidungen und Zeitverwendung bei Paaren" (Accompanying documentation for the study "Consumption decisions and time-use of couples"). Research project documentation, Berlin School of Economics.

Beblo, M. and D. Beninger (2012). Do Husbands and Wives Pool Their Incomes? Experimental Evidence. Discussion paper, Bureau d'Economie Théorique et Appliquée, UDS, Strasbourg.

Beblo, M. and D. Beninger (2017). Do husbands and wives pool their incomes? A couple experiment. *Review of Economics of the Household 15*(3), 779–805.

Beblo, M., D. Beninger, F. Cochard, H. Couprie, and A. Hopfensitz (2015). Efficiency-Equality Trade-off within French and German couples: A comparative experimental study. *Annals of Economics and Statistics/Annales d'Économie et de Statistique* (117-118), 233–252.

Beblo, M., A. Heinze, and E. Wolf (2008). Entwicklung der beruflichen Segregation von Männern und Frauen zwischen 1996 und 2005. Eine Bestandsaufnahme auf betrieblicher Ebene (Occupational segregation of men and women between 1996 and 2005 – an analysis at the establishment level). *Zeitschrift für ArbeitsmarktForschung 41*(2), 3.

Beblo, M. and J. R. Robledo (2008). The wage gap and the leisure gap for double-earner couples. *Journal of Population Economics 21*(2), 281–304.

Becker, G. S. (1957). *The theory of discrimination*. University of Chicago Press.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, 493–517.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, 813–846.

Becker, G. S. (1974a). A theory of marriage: Part II. *Journal of Political Economy 82*(2, Part 2), S11–S26.

Becker, G. S. (1974b). A theory of social interactions. *Journal of Political Economy 82*(6), 1063–1093.

Becker, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics 3*(1, Part 2), S33–S58.

Becker, G. S. (1991). *A Treatise on the Family*. Harvard University Press.

Becker, G. S. and H. G. Lewis (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy 81*(2, Part 2), S279–S288.

Becker, S. O., F. Cinnirella, E. Hornung, and L. Woessmann (2012). iPEHD: The ifo Prussian Economic History Database. Working paper, CESifo: Economics of Education.

Becker, S. O., F. Cinnirella, and L. Woessmann (2013). Does women's education affect fertility? Evidence from pre-demographic transition Prussia. *European Review of Economic History 17*(1), 24–44.

Becker, S. O. and L. Woessmann (2008). Luther and the girls: Religious denomination and the female education gap in nineteenth-century Prussia. *The Scandinavian Journal of Economics 110*(4), 777–805.

Beit-Hallahmi, B., J. C. Catford, R. E. Cooley, C. Y. Dull, A. Z. Guiora, and M. Raluszny (1974). Grammatical gender and gender identity development: Cross cultural and cross lingual implications. *American Journal of Orthopsychiatry 44*(3), 424.

Beller, S., K. F. Brattebø, K. O. Lavik, R. D. Reigstad, and A. Bender (2015). Culture or language: what drives effects of grammatical gender? *Cognitive Linguistics 26*(2), 331–359.

Ben-Ner, A., F. Kong, and L. Putterman (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *Journal of Economic Psychology 25*(5), 581–589.

Bénabou, R. and J. Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics 126*(2), 805–855.

Benndorf, V., H. A. Rau, and C. Sölch (2014). Minimizing learning behavior in experiments with repeated real-effort tasks. Working paper, SSRN. Available online: `http://dx.doi.org/10.2139/ssrn.2503029` (accessed May 18, 2018).

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*(1), 122–142.

Bertrand, M. (2011). New perspectives on gender. In O. Ashenfelter and D. Card (Eds.), *Handbook of labor economics*, Volume 4, pp. 1543–1590. Elsevier.

Bertrand, M., P. Cortés, C. Olivetti, and J. Pan (2018). Social norms, labor market opportunities, and the marriage gap for skilled women. IZA discussion paper no.11382.

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

Beveridge, F., F. Cengiz, E. Drywood, M. Corsi, D. Szelewa, E. Lépinard, L. Altan, and P. Debusscher (2016). A new strategy for gender equality post 2015. Technical report, European Parliament. `http://www.europarl.europa.eu/RegData/etudes/STUD/2014/509984/IPOL_STU(2014)509984_EN.pdf` (accessed May 18, 2018).

Bittman, M., P. England, L. Sayer, N. Folbre, and G. Matheson (2003). When Does Gender Trump Money? Bargaining and Time in Household Work. *American Journal of Sociology 109*(1), pp. 186–214.

Black, D. A., S. G. Sanders, and L. J. Taylor (2007). The economics of lesbian and gay families. *Journal of Economic Perspectives 21*(2), 53–70.

Blau, F. D. and L. M. Kahn (2013). Female labor supply: Why is the united states falling behind? *American Economic Review 103*(3), 251–56.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Blundell, R., P.-A. Chiappori, and C. Meghir (2005). Collective labor supply with children. *Journal of political Economy 113*(6), 1277–1306.

BMAS (Federal Ministry for Labor and Social Affairs Germany) (2006). *Statistische Übersichten zur Sozialpolitik in Deutschland seit 1945, Band SBZ/DDR*. Online: `http://www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/fb-f352-statistische-uebersichten-zur-sozialpolitik-in-deutschland-/seit-1945.pdf` (accessed May 18, 2018).

Bobonis, G. J., M. González-Brenes, and R. Castro (2013). Public transfers and domestic violence: The roles of private information and spousal control. *American Economic Journal: Economic Policy*, 179–205.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review 71*, 117–120.

Booth, A. L. and P. Nolen (2012a). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization 81*(2), 542–555.

Booth, A. L. and P. Nolen (2012b). Gender differences in risk behaviour: Does nurture matter? *The Economic Journal 122*(558), F56–F78.

Borck, R. (2014). Adieu rabenmutter—culture, fertility, female labour supply, the gender wage gap and childcare. *Journal of Population Economics 27*(3), 739–765.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Boroditsky, L., L. A. Schmidt, and W. Phillips (2003). Sex, syntax, and semantics. In D. Genter and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*, pp. 61–79. MIT.

Bothfeld, S. (2005). Arbeitsmarkt (Labour market). In S. Bothfeld, U. Klammer, C. Klenner, S. Leiber, A. Thiel, and A. Ziegler (Eds.), *Frauendatenreport 2005. Handbuch zur wirtschaftlichen und sozialen Situation von Frauen (Women data report. Handbook on the economic and social situation of women)*, pp. 109–186.

Braun, M., J. Scott, and D. F. Alwin (1994). Economic necessity or self-actualization? Attitudes toward women's labour-force participation in East and West Germany. *European Sociological Review 10*(1), 29–47.

Browning, M., P.-A. Chiappori, and Y. Weiss (2014). *Economics of the Family*. Cambridge University Press.

Bühlmann, F., G. Elcheroth, and M. Tettamanti (2010). The division of labour among European couples: The effects of life course and welfare policy on value-practice configurations. *European Sociological Review 26*(1), 49–66.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017). 'Acting Wife': Marriage market incentives and labor market investments. *American Economic Review 107*(11), 3288–3319.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, Competitiveness, and career choices. *The Quarterly Journal of Economics 1409*, 1447.

Camerer, C. and R. H. Thaler (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives 9*(2), 209–219.

Campa, P. and M. Serafinelli (2016). Politico-economic regimes and attitudes: Female workers under state-socialism. Unpublished Manuscript, University of Toronto, available online: `https://sites.google.com/site/michelserafinelli/` (accessed May 18, 2018).

Cárdenas, J.-C., A. Dreber, E. von Essen, and E. Ranehill (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior & Organization 83*(1), 11–23.

Carlsson, F., P. Martinsson, P. Qin, and M. Sutter (2013). The influence of spouses on household decision making under risk: An experiment in rural China. *Experimental Economics 16*(3), 383–401.

Casasanto, D. (2016). Linguistic relativity. In N. Riemer (Ed.), *The Routledge handbook of semantics*, pp. 158–174. Routledge.

Casasanto, D., L. Boroditsky, W. Phillips, J. Greene, S. Goswami, S. Bocanegra-Thiel, I. Santiago-Diaz, O. Fotokopoulu, R. Pita, and D. Gil (2004). How deep are effects of language on thought? time estimation in speakers of english, indonesian, greek, and spanish. In *Proceedings of the Cognitive Science Society*, Volume 26.

Castilla, C. and T. Walker (2013). Is ignorance bliss? The effect of asymmetric information between spouses on intra-household allocations. *American Economic Review 103*(3), 263–268.

Charness, G. and U. Gneezy (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization 83*(1), 50–58.

Charness, G. and P. Kuhn (2011). Lab labor: What can labor economists learn from the lab? Volume 4 of *Handbook of Labor Economics*, pp. 229 – 330. Elsevier.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 817–869.

Chen, M. K. (2013, April). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review 103*(2), 690–731.

Chen, Z. and F. Woolley (2001). A cournot–nash model of family decision making. *The Economic Journal 111*(474), 722–748.

Chevalier, A. and O. Marie (2017). Economic uncertainty, parental selection, and children's educational outcomes. *Journal of Political Economy 125*(2), 393–430.

Chiappori, P.-A. (1988). Rational household labor supply. *Econometrica: Journal of the Econometric Society 56(1)*, 63–90.

Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of Political Economy 100*(3), 437–467.

Chiswick, B. R. (1999). Are immigrants favorably self-selected? *American Economic Review 89*(2), 181–185.

Chiswick, B. R. and P. W. Miller (2015). International migration and the economics of language. In B. R. Chiswick and P. W. Miller (Eds.), *Handbook of the economics of international migration*, Volume 1, pp. 211–269. Elsevier.

Cobb-Clark, D. A. and M. Tan (2011). Noncognitive skills, occupational attainment, and relative wages. *Labour Economics 18*(1), 1–13.

Cochard, F., H. Couprie, and A. Hopfensitz (2009). Do Spouses Cooperate? And If Not: Why? Discussion paper, THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

Cochard, F., H. Couprie, and A. Hopfensitz (2013). What if women earned more than their spouse? Conference Paper, XXXèmes Journées de Microéconomie Appliquée.

Cochard, F., H. Couprie, and A. Hopfensitz (2014). Do spouses cooperate? An experimental investigation. *Review of Economics of the Household*, 1–26.

Cochard, F., H. Couprie, and A. Hopfensitz (2018, Mar). What if women earned more than their spouses? An experimental investigation of work-division in couples. *Experimental Economics 21*(1), 50–71.

Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (2018, January). The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers. Research Paper repec:ecl:stabus:3637, Stanford University, Graduate School of Business.

Cooke, L. P. (2006). Policy, preferences, and patriarchy: The division of domestic labor in East Germany, West Germany, and the United States. *Social Politics: International Studies in Gender, State & Society 13*(1), 117–143.

Corbett, G. G. (2013a). Number of genders. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Corbett, G. G. (2013b). Sex-based and non-sex-based gender systems. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Corbett, G. G. (2013c). Systems of gender assignment. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Crawford, V. P., U. Gneezy, and Y. Rottenstreich (2008). The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. *American Economic Review 98*(4), 1443–58.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature*, 448–474.

Cudeville, E. and M. Recoules (2015). Household behaviour and social norms: A conjugal contract model with conformism. *Annals of Economics and Statistics/Annales d'Économie et de Statistique* (117-118), 279–312.

Dahl, D. B. (2016). xtable: Export tables to latex or html. `https://CRAN.R-project.org/package=xtable` (accessed May 18, 2018). R package version 1.8-2.

Dahlkamp, S., I. Hülsen, A.-K. Müller, and A. Seith (2017). Reden ist Geld. (talk is money). *Der Spiegel* (3).

Dato, S. and P. Nieken (2014). Gender differences in competition and sabotage. *Journal of Economic Behavior & Organization 100*, 64–80.

David, H. P. (1992). Abortion in Europe, 1920-91: A public health perspective. *Studies in Family Planning 23*(1), 1–22.

Destatis (2015). On-demand data provision by the Federal Statistical Office of Germany, Wiesbaden.

Destatis (Federal Statistical Office Germany) (1972). *Bevölkerung und Wirtschaft 1872-1972 (Population and Economy 1872-1972)*. Online: `https://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEMonografie_derivate_00000198/Bevoelkerung_und_%20Wirtschaft%201872_1972.pdf` (accessed May 18, 2018).

Destatis (Federal Statistical Office Germany) (1982). *Statistisches Jahrbuch (Statistical Yearbook)*. Online: `http://www.digizeitschriften.de/dms/toc/?PID=PPN514402342_1990` (accessed May 18, 2018).

Destatis (Federal Statistical Office Germany) (1990). *Statistisches Jahrbuch (Statistical Yearbook)*. Online: `http://www.digizeitschriften.de/dms/toc/?PID=PPN514402342_1990` (accessed May 18, 2018).

Destatis (Federal Statistical Office Germany) (2014a). *Bevölkerung und Erwerbstätigkeit. Stand und Entwicklung der Erwerbstätigkeit in Deutschland (Population and employment. Level and trends in employment in Germany)*. Online: `https://www.destatis.de/DE/Publikationen/Thematisch/Arbeitsmarkt/`

Erwerbstaetige/StandEntwicklungErwerbstaetigkeit2010411157004.pdf?__blob=publicationFile (accessed May 18, 2018).

Destatis (Federal Statistical Office Germany) (2014b). *Gender Pay Gap.* Online: https://www.destatis.de/EN/FactsFigures/Indicators/QualityEmployment/Dimension1.html (accessed May 18, 2018).

Destatis (Federal Statistical Office Germany) (2016). *Bevölkerung und Erwerbstätigkeit. Zusammenfassende Übersichten Eheschließungen, Geborene und Gestorbene (Population and employment. Summary of marriages, births and deaths).* Online: https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/ZusammenEheschliessungenGeboreneGestorbene5126102157004.pdf?__blob=publicationFile (accessed May 18, 2018).

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–590.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Donni, O. and P.-A. Chiappori (2011). Nonunitary models of household behavior: A survey of the literature. In J. A. Molina (Ed.), *Household economic behaviors*, pp. 1–40. Springer.

Dragulescu, A. A. (2014). xlsx: Read, write, format excel 2007 and excel 97/2000/xp/2003 files. https://CRAN.R-project.org/package=xlsx. R package version 0.5.7.

Dreber, A., E. von Essen, and E. Ranehill (2011). Outrunning the gender gap—boys and girls compete equally. *Experimental Economics 14*(4), 567–582.

Dryer, M. S. and M. Haspelmath (Eds.) (2013). *WALS Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa. *The World Bank Economic Review 17*(1), 1–25.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature 50*(4), 1051–1079.

Duggan, L. (1995). Restacking the deck: Family policy and women's fall-back position in Germany before and after unification. *Feminist Economics 1*(1), 175–194.

Eagly, A. H. and W. Wood (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist 54*(6), 408.

Eckel, C. C. and P. J. Grossman (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal 108*(448), 726–735.

EIGE (2017a). Gender Equality Index 2017 in brief. A snail's pace towards gender equality. European institute for gender equality, European Institute for Gender Equality publication. Available online: `http://eige.europa.eu/rdc/eige-publications/gender-equality-index-2017-brief-snails-pace-towards-gender-equality` (accessed May 18, 2018).

EIGE (2017b). Gender Equality Index 2017. Measuring gender equality in the European Union 2005-2015. Policy report. Available online: `http://eige.europa.eu/rdc/eige-publications/gender-equality-index-2017-measuring-gender-equality-european\-union-2005-2015-report`.

Ekerdt, D. J., K. Kosloski, and S. DeViney (2000). The normative anticipation of retirement by older workers. *Research on Aging 22*(1), 3–22.

Engineer, M. and L. Welling (1999). Human capital, true love, and gender roles: is sex destiny? *Journal of Economic Behavior & Organization 40*(2), 155–178.

Erkal, N., L. Gangadharan, and N. Nikiforakis (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review 101*(7), 3330–3348.

Esping-Andersen, G. (2009). *Incomplete revolution: Adapting welfare states to women's new roles*. Cambridge, Polity Press.

European Commission (2015a). Report on equality between women and men 2010. Commission staff working document, Accompanying document to the report from the Commission to the Coun-

cil, the EP, the EESC and the CoR, COM (2015). Available online: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0046+0+DOC+XML+V0//EN (accessed May 18, 2018).

European Commission (2015b). Report on equality between women and men 2015. Commission s taff working document, Accompanying document to the report from the Commission to the Council, the EP, the EESC and the CoR, COM (2015). Available online: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0046+0+DOC+XML+V0//EN (accessed May 18, 2018).

European Commission (2016). Strategic engagement for gender equality 2016-2019. Technical report. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/gender-equality/gender-equality-strategy_en (accessed May 18, 2018).

Eurostat (2013a). Table: Employment rate by sex, age groups, highest level of education attained and household composition (%). Dataset-code: lfst_hheredty. ec.europa.eu/eurostat. http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database (accessed October 23, 2013).

Eurostat (2013b). Table: Inactive population - Main reason for not seeking employment - Distributions by sex and age (%). Dataset-code: lfsa_igar. ec.europa.eu/eurostat. http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_lfs/data/database (accessed October 23, 2013).

Eurostat (2013c). Table: Main reason for part-time employment – Distributions by sex and age (%). Dataset-code: lfsa_epgar. ec.europa.eu/eurostat. http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_lfs/data/database (accessed October 23, 2013).

Eurostat (2017). Table: Employment rate by sex, age groups, highest level of education attained and household composition (%). Dataset-code: lfst_hheredty. ec.europa.eu/eurostat. http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database (accessed May 18, 2018).

Fagan, C. and H. Norman (2013). Men and gender equality. Tackling gender segregated family roles and social care jobs. In *Commission staff working document*, Volume 77, pp. 199–223. Routledge.

Fahn, M., R. Rees, and A. Wuppermann (2016). Relational contracts for household formation, fertility choice and separation. *Journal of Population Economics 29*(2), 421–455.

Falk, A., E. Fehr, and U. Fischbacher (2008). Testing theories of fairness – Intentions matter. *Games and Economic Behavior 62*(1), 287–303.

Farré, L. and F. Vella (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica 80*(318), 219–247.

Fassmann, H. and R. Munz (1994). European East-West migration, 1945-1992. *International Migration Review*, 520–538.

Fehr, E. and S. Gächter (2000a). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

Fehr, E. and S. Gächter (2000b). Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives 14*(3), 159–181.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review 103*(1), 472–500.

Fernandez, R. and A. Fogli (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics 1*(1), 146–177.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, 1249–1299.

Fernández, R. (2011). Does culture matter? In J. M. Benhabib, J. and A. Bisin (Eds.), *Handbook of Social Economics*, Volume Vol. 1A. North-Holland.

Filippin, A. and P. Crosetto (2016). A reconsideration of gender differences in risk attitudes. *Management Science 62*(11), 3138–3160.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics 121*(2), 673–697.

Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology 32*(1), 18–31.

Fogli, A. and L. Veldkamp (2011). Nature or nurture? Learning and the geography of female labor force participation. *Econometrica 79*(4), 1103–1138.

Ford, A. C. (2017). Millennial women are conflicted about being breadwinners. Available online: `https://www.refinery29.com/2017/04/148488/millennial-women-are-conflicted-about-being-breadwinners` (accessed May 18, 2018).

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.

Fortin, N. M. (2008). The gender wage gap among young adults in the united states the importance of money versus people. *Journal of Human Resources 43*(4), 884–918.

Fortin, N. M. (2015). Gender role attitudes and women's labor market participation: Opting-out, aids, and the persistent appeal of housewifery. *Annals of Economics and Statistics/Annales d'Économie et de Statistique* (117-118), 379–401.

Fox, J. and S. Weisberg (2011). *An R Companion to Applied Regression* (Second ed.). Thousand Oaks CA: Sage.

Friehe, T. and M. Mechtel (2014). Conspicuous consumption and political regimes: Evidence from East and West Germany. *European Economic Review 67*, 62–81.

Fuchs-Schündeln, N. and P. Masella (2016). The long lasting effects of socialist education. *The Review of Economics and Statistics 98*(3), 428–441.

Galor, O., Ö. Özak, and A. Sarid (2017). Geographical origins and economic consequences of language structures. Ssrn working paper. Available online: `https://ssrn.com/abstract=2820889orhttp://dx.doi.org/10.2139/ssrn.2820889` (accessed May 18, 2018).

Gay, V., D. L. Hicks, E. Santacreu-Vasut, and A. Shoham (2017, Apr). Decomposing culture: an analysis of gender, language, and labor supply in the household. *Review of Economics of the Household*.

Gay, V., E. Santacreu-Vasut, and A. Shoham (2013). The grammatical origins of gender roles. Working paper, Berkeley Economic History Lab. Available online: `http://behl.berkeley.edu/working-papers/` (accessed May 18, 2018).

Gerhard, U. (1992). German women and the social costs of unification. *German Politics and Society 24*(25), 1991–92.

Gerhards, L. and N. Siemer (2016). The impact of private and public feedback on worker performance–evidence from the lab. *Economic Inquiry 54*(2), 1188–1201.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Gill, D. and V. Prowse (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics 5*(2), 351–376.

Givati, Y. and U. Troiano (2012). Law, economics, and culture: Theory of mandated benefits and evidence from maternity leave policies. *The Journal of Law and Economics 55*(2), 339–364.

Gneezy, U., K. L. Leonard, and J. A. List (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica 77*(5), 1637–1664.

Goldin, C. (1995). The U-shaped female labor force function in economic development and economic history. In T. Schultz (Ed.), *Investment in women's human capital and economic development*, pp. 61–90. University of Chicago Press.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review 104*(4), 1091–1119.

Goldin, C., S. P. Kerr, C. Olivetti, and E. Barth (2017). The expanding gender earnings gap: Evidence from the LEHD-2000 census. *American Economic Review: Papers and Proceedings 107*(5), 110–114.

Gordon, R. G. (Ed.) (2005). *Ethnologue, languages of the world.* sil International. `http://www.ethnologue.com/` (accessed May 18, 2018).

Görges, L. (2015). The power of love: A subtle driving force for unegalitarian labor division? *Review of Economics of the Household 13*(1), 163–192.

Graham, N., M. Arai, and B. Hagströmer (2016). multiwayvcov: Multi-way standard error clustering. `https://CRAN.R-project.org/package=multiwayvcov` (accessed May 18, 2018). R package version 1.2.3.

Greenwood, J., A. Seshadri, and G. Vandenbroucke (2005). The baby boom and baby bust. *American Economic Review 95*(1), 183–207.

Greenwood, J., A. Seshadri, and M. Yorukoglu (2005). Engines of liberation. *The Review of Economic Studies 72*(1), 109–133.

Gronau, R. (1973a). The effect of children on the housewife's value of time. *Journal of Political Economy 81*(2, Part 2), S168–S199.

Gronau, R. (1973b). The intrafamily allocation of time: The value of the housewives' time. *American Economic Review 63*(4), 634–651.

Gronau, R. (1977). Leisure time, home production, and market work. *Journal of Political Economy 85*, 1099–1124.

Grossbard, S. (2011). Independent individual decision-makers in household models and the new home economics. In J. A. Molina (Ed.), *Household economic behaviors*, pp. 41–56. Springer.

Grove, W. A., A. Hussey, and M. Jetter (2011). The gender pay gap beyond human capital heterogeneity in noncognitive skills and in labor market tastes. *Journal of Human Resources 46*(4), 827–874.

Groves, M. O. (2005). How important is your personality? Labor market returns to personality for women in the US and UK. *Journal of Economic Psychology 26*(6), 827–841.

Gubler, J. (2014). Prediction function for linear regression with robust/cluster robust standard errors. Available online: `https://gist.github.com/13bzhang/98ba16677aca813da60c3173cf839562#file-predict_lm-r` (accessed May 18, 2018).

Guiora, A. Z., B. Beit-Hallahmi, R. Fried, and C. Yoder (1982). Language environment and gender identity attainment. *Language Learning 32*(2), 289–304.

Günther, C., N. A. Ekinci, C. Schwieren, and M. Strobel (2010). Women can't jump? – An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization 75*(3), 395–401.

Güth, W., R. Ivanova-Stenzel, M. Sutter, H. Weck-Hannemann, et al. (2003). Investment and bargaining in joint ventures: A family decision making experiment. *Journal of Institutional and Theoretical Economics (JITE) 159*(2), 323–323.

Güth, W., R. Ivanova-Stenzel, and S. Tjotta (2004). Please, marry me! An experimental study of risking a joint venture. *Metroeconomica 55*(1), 1–21.

Haberkern, K. (2007). Zeitverwendung und Arbeitsteilung in Paarhaushalten (Time-use and labor division in couples' households). *Zeitschrift für Familienforschung 19*(2), 159–185.

Hadfield, G. K. (1999). A coordination model of the sexual division of labor. *Journal of Economic Behavior & Organization 40*(2), 125–153.

Haertlé, I. (2017). Does grammatical gender influence perception? A study of Polish and French speakers. *Psychology of Language and Communication 21*(1), 386–407.

Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic Literature 42*(4), 1009–1055.

Hart, O. (1995). *Firms, contracts, and financial structure*. Clarendon Press.

Hart, O. and J. Moore (1988). Incomplete contracts and renegotiation. *Econometrica: Journal of the Econometric Society*, 755–785.

Hart, O. and J. Moore (1990). Property rights and the nature of the firm. *Journal of Political Economy 98*(6), 1119–1158.

Hechavarría, D. M., S. A. Terjesen, P. Stenholm, M. Brännback, and S. Lång (2017). More than words: do gendered linguistic structures widen the gender gap in entrepreneurial activity? *Entrepreneurship Theory and Practice*.

Hederos Eriksson, K. and A. Stenberg (2015). Gender identity and relative income within households: Evidence from sweden. Discussion Paper 9533, Institute for the Study of Labor (IZA).

Helson, R. and C. J. Soto (2005). Up and down in middle age: Monotonic and nonmonotonic changes in roles, status, and personality. *Journal of Personality and Social Psychology 89*(2), 194–204.

Hicks, D. L., E. Santacreu-Vasut, and A. Shoham (2015). Does mother tongue make for women's work? Linguistics, household labor, and gender identity. *Journal of Economic Behavior & Organization 110*, 19–44.

Hlavac, M. (2018). stargazer: Well-formatted regression and summary statistics tables. `https://CRAN.R-project.org/package=stargazer` (accessed May 18, 2018). R package version 5.2.1.

Hoel, J. B. (2015). Heterogeneous households: A within-subject test of asymmetric information between spouses in Kenya. *Journal of Economic Behavior & Organization 118*, 123–135.

Holm, H. J. (2000). Gender-based focal points. *Games and Economic Behavior 32*(2), 292–314.

Holst, E. and J. Schupp (2001). Erwerbsverhalten von Frauen: trotz Annäherung immer noch deutliche Unterschiede zwischen Ost und West (Women's employment: Despite convergence still substantial differences between East and West). DIW weekly report 42.

Hothorn, T., F. Bretz, and P. Westfall (2008). Simultaneous inference in general parametric models. *Biometrical Journal 50*(3), 346–363.

Humlum, M. K., K. J. Kleinjans, and H. S. Nielsen (2012). An economic analysis of identity and career choice. *Economic Inquiry 50*(1), 39–61.

Hunt, E. and F. Agnoli (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review 98*(3), 377.

Hunt, J. (2002). The transition in East Germany: When Is a ten-point fall in the gender wage gap bad news? *Journal of Labor Economics 20*(1), 148–169.

Hunt, J. (2006). Staunching emigration from East Germany: Age and the determinants of migration. *Journal of the European Economic Association 4*(5), 1014–1037.

Hyll, W. and L. Schneider (2013). The causal effect of watching TV on material aspirations: Evidence from the "valley of the innocent". *Journal of Economic Behavior & Organization 86*, 37–51.

ILO (2018). ILOSTAT: Labour force participation rate by sex and age (%). http://www.ilo.org/ilostat/faces/oracle/webcenter/portalapp/pagehierarchy/Page27.jspx?subject=EAP&indicator=EAP_DWAP_SEX_AGE_RT&datasetCode=A&collectionCode=YI&_afrLoop=633361233136369&_afrWindowMode=0&_afrWindowId=t4u2a38my_1#!%40%40%3Findicator%3DEAP_DWAP_SEX_AGE_RT%26_afrWindowId%3Dt4u2a38my_1%26subject%3DEAP%26_afrLoop%3D633361233136369%26datasetCode%3DA%26collectionCode%3DYI%26_afrWindowMode%3D0%26_adf.ctrl-state%3Dt4u2a38my_66 (accessed May 18, 2018).

Isoni, A., A. Poulsen, R. Sugden, and K. Tsutsui (2013). Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review 59*, 167–188.

Isoni, A., A. Poulsen, R. Sugden, and K. Tsutsui (2014). Efficiency, equality, and labeling: An experimental investigation of focal points in explicit bargaining. *American Economic Review 104* (10), 3256–3287.

Iversen, V., C. Jackson, B. Kebede, A. Munro, and A. Verschoor (2011). Do spouses realise cooperative gains? Experimental evidence from rural Uganda. *World Development 39* (4), 569–578.

Judge, T. A. and C. Hurst (2007). Capitalizing on one's advantages: role of core self-evaluations. *Journal of Applied Psychology 92* (5), 1212.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Kalleberg, A. L. and K. A. Loscocco (1983). Aging, values, and rewards: Explaining age differences in job satisfaction. *American Sociological Review*, 78–90.

Kebede, B., M. Tarazona, A. Munro, and A. Verschoor (2013). Intra-household efficiency: An experimental study from Ethiopia. *Journal of African Economies*, ejt019.

Kleven, H., C. Landais, and J. E. Søgaard (2018). Children and gender inequality: Evidence from denmark. Working paper 24219, National Bureau of Economic Research.

Klüsener, S. and J. R. Goldstein (2014). A long-standing demographic East–West divide in Germany. *Population, Space and Place*. DOI: 10.1002/psp.1870.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of psycholinguistic research 22*(5), 519–534.

Konrad, K. A. and K. E. Lommerud (1995). Family policy with non-cooperative families. *The Scandinavian Journal of Economics 97*(4), 581–601.

Konrad, K. A. and K. E. Lommerud (2000). The bargaining family revisited. *Canadian Journal of Economics/Revue canadienne d'économique 33*(2), 471–487.

Krack-Roberg, E., S. Krieger, and J. Weinmann (2016). Familie, Lebensformen und Kinder (Family, lifestyles, and children). In Destatis (Ed.), *Datenreport (Data report)*, pp. 43–77.

Kröhnert, S. and S. Skipper (2010). Demographic development in Eastern Germany. *Online Handbook Demography. Berlin Institute for Population and Development*. Online: `https://www.berlin-institut.org/en/online-handbookdemography/population-dynamics/east-germany.html` (accessed May 18, 2018).

Krueger, A. B. and J.-S. Pischke (1995). A comparative ananlysis of East and West German labor markets: Before and after unification. In R. B. Freeman and L. F. Katz (Eds.), *Differences and Changes in Wage Structures*, pp. 405–446.

Kuhn, P. and K. Shen (2013). Gender discrimination in job ads: Evidence from china. *The Quarterly Journal of Economics 128*(1), 287–336.

Lazear, E. P. and S. Rosen (1981). Rank-Order tournaments as optimum labor contracts. *Journal of Political Economy*, 841–864.

Lazear, E. P. and S. Rosen (1990). Male-female wage differentials in job ladders. *Journal of Labor Economics*, S106–S123.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software 69*(1), 1–33.

Lück, D. (2015). Vaterleitbilder: Ernährer und Erzieher. In N. Schneider, S. Diabaté, and K. Ruckdeschel (Eds.), *Familienleitbilder in Deutschland. Kulturelle Vorstellungen zu Partnerschaft, Elternschaft und Familienleben. Beiträge zur Bevölkerungswissenschaft*, Volume 48, pp. 227–245. Verlag Barbara Budrich.

Lundberg, S. and R. A. Pollak (1993). Separate spheres bargaining and the marriage market. *Journal of Political Economy 101*(6), 988–1010.

Lundberg, S. and R. A. Pollak (1994). Noncooperative bargaining models of marriage. *American Economic Review 84*(2), 132–137.

Lundberg, S., R. A. Pollak, and J. Stearns (2016). Family inequality: Diverging patterns in marriage, cohabitation, and childbearing. *The Journal of Economic Perspectives 30*(2), 79–101.

Lundberg, S. J., R. A. Pollak, and T. J. Wales (1997). Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human resources*, 463–480.

Lundborg, P., E. Plug, and A. W. Rasmussen (2017). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review 107*(6), 1611–37.

MacKay, D. G. (1999). *Gender in English, German, and other languages: Problems with the old theory, opportunities for the new*, pp. 73–87. Wiesbaden: VS Verlag für Sozialwissenschaften.

Maier, F. (1993). The labour market for women and employment perspectives in the aftermath of German unification. *Cambridge Journal of Economics 17*(3), 267–280.

Mani, A. Mine, yours or ours? The efficiency of household investment decisions: An experimental approach. CAGE online working paper, University of Warwick, Department of Economics.

Manser, M. and M. Brown (1980). Marriage and household decision-making: A bargaining analysis. *International Economic Review*, 31–44.

Mavisakalyan, A. (2015). Gender in language and gender in employment. *Oxford Development Studies 43*(4), 403–424.

Mavisakalyan, A. and C. Weber (2017). Linguistic structures and economic outcomes. *Journal of Economic Surveys*.

McElroy, M. B. and M. J. Horney (1981). Nash-bargained household decisions: Toward a generalization of the theory of demand. *International Economic Review 22*(2), 333–349.

Meyers Konversations-Lexikon (1896). *Preußen* (5 ed.), Volume 14. Online: `http://www.digizeitschriften.de/main=/dms/img/?PPN=PPN514401303_1935` (accessed May 18, 2018).

Miller, A. and S. Sassler (2010). Stability and change in the division of labor among cohabiting couples. *25*(4), 677–702.

Millo, G. (2017). Robust standard error estimators for panel models: A unifying approach. *Journal of Statistical Software 82*(3), 1–27.

Moffatt, P. G. (2015). *Experimetrics: Econometrics for experimental economics*. Palgrave Macmillan.

Munro, A., I. Bateman, and T. McNally (2008). The family under the microscope: an experiment testing economic models of household choice. Munich personal repec archive discussion paper no. 8974.

Nelson, J. A. (2015). Are women really more risk-averse than men? a re-analysis of the literature using expanded methods. *Journal of Economic Surveys 29*(3), 566–585.

Niederle, M. (2016). Gender. In J. H. Kagel and A. E. Roth (Eds.), *The Handbook of Experimental Economics*.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annu. Rev. Econ. 3*(1), 601–630.

OECD (2012). Gender Equality in Education, Employment and Entrepreneurship: Final Report to the MCM 2012. Technical report. Available online: http://www.oecd.org/social/familiesandchildren/50423364.pdf.

OECD (2016). Table: Employment rate (indicator). Available online: `https://data.oecd.org/emp/employment-rate.htm` (accessed May 18, 2018).

OECD (2017a). Report on the implementation of the OECD gender recommendations - Some progress on gender equality but much left to do. Technical report. Available online: `http://www.oecd.org/mcm/documents/C-MIN-2017-7-EN.pdf` (accessed May 18, 2018).

OECD (2017b). *The pursuit of gender equality. An uphill battle.* Online: https://www.oecd-ilibrary.org/content/publication/9789264281318-en (accessed May 18, 2018).

OECD family database (2016). Table lmf2.5.a: Time allocated in main activities, by gender, 1999-2012/2013a. http://www.oecd.org/els/family/database.htm (accessed May 18, 2018).

Oosterbeek, H., J. Sonnemans, and S. Van Velzen (2003). The need for marriage contracts: An experimental study. *Journal of Population Economics 16*(3), 431–453.

Ostner, I. (1993). Slow motion: women, work and the family in Germany. In J. Lewis (Ed.), *Women and social policies in Europe: Work, family and the state*, pp. 110–12. Edward Elgar Publishing Ltd.

Pedulla, D. S. and S. Thébaud (2015). Can we finish the revolution? Gender, work-family ideals, and institutional constraint. *American Sociological Review 80*(1), 116–139.

Peters, H. E., A. S. Ünür, J. Clark, and W. D. Schulze (2004). Free-riding and the provision of public goods in the family: A laboratory experiment. *International Economic Review 45*(1), 283–299.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review 62*(4), 659–661.

Phipps, S. A. and P. S. Burton (1998). What's mine is yours? the influence of male and female incomes on patterns of household expenditure. *Economica 65*(260), 599–613.

Pollak, R. A. and M. L. Wachter (1975). The relevance of the household production function and its implications for the allocation of time. *Journal of Political Economy 83*(2), 255–278.

Quisumbing, A. R. and J. A. Maluccio (2003). Resources at marriage and intrahousehold allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa. *Oxford Bulletin of Economics and Statistics 65*(3), 283–327.

R Core Team (2018). R: A language and environment for statistical computing. https://www.R-project.org/ (accessed May 18, 2018).

Rainer, H. and T. Siedler (2009). Does democracy foster trust? *Journal of Comparative Economics 37*(2), 251–269.

Reid, M. G. (1934). *Economics of household production*. Wiley.

Reuben, E., M. Wiswall, and B. Zafar (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2168.

Roberts, S. and J. Winters (2013). Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PloS one 8*(8), e70902.

Roberts, S. G., J. Winters, and K. Chen (2015). Future tense and economic decisions: controlling for cultural evolution. *PloS one 10*(7), e0132145.

Robertson, J. A. (2004). Reproductive technology in Germany and the United States: An essay in comparative law and bioethics. *Colum. J. Transnat'l L. 43*, 189.

Robinson, D. (2018). broom: Convert statistical analysis objects into tidy data frames. `https://CRAN.R-project.org/package=broom` (accessed May 18, 2018). R package version 0.4.4.

Rosenfeld, R. A. and H. Trappe (2002). Occupational sex segregation in state socialist and market economies: Levels, patterns, and change in East and West Germany, 1980s and 1998. *Research in Social Stratification and Mobility 19*, 231–267.

Rosenfeld, R. A., H. Trappe, and J. C. Gornick (2004). Gender and work in Germany: Before and after reunification. *Annual Review of Sociology 30*, 103–124.

Samuelson, P. A. (1956). Social indifference curves. *The Quarterly Journal of Economics 70*(1), 1–22.

Santacreu-Vasut, E., O. Shenkar, and A. Shoham (2014). Linguistic gender marking and its international business ramifications. *Journal of International Business Studies 45*(9), 1170–1178.

Santacreu-Vasut, E., A. Shoham, and V. Gay (2013). Do female/male distinctions in language matter? Evidence from gender political quotas. *Applied Economics Letters 20*(5), 495–498.

Saraceno, C. (2011). Gender (in) equality: An incomplete revolution? Cross EU similarities and differences in the gender specific impact of parenthood. Discussion Paper 13(3), Harriet Taylor Mill-Institut, Berlin School of Economics and Law.

Säve-Söderbergh, J. and G. S. Lindquist (2017). Children do not behave like adults: Gender gaps in performance and risktaking in a random social context in the high-stakes game shows Jeopardy and Junior Jeopardy. *The Economic Journal 127*(603), 1665–1692.

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.

Schmitt, N. (2013). An Experiment on the Influence of Stereotypes on Economic Decision-Making. Conference Paper, 13th annual EALE conference.

Schröder, M., N. Schmitt, B. Heynemann, and C. Brünn (2013). Income Taxation and Labor Supply: An Experiment on Couple's Work Effort. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013. Available online: `http://hdl.handle.net/10419/79735` (accessed May 18, 2018).

Sen, A. (2001). The many faces of gender inequality. *New Republic*, 35–39.

Sera, M. D., C. A. Berge, and J. del Castillo Pintado (1994). Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers. *Cognitive Development 9*(3), 261–292.

Sera, M. D., C. Elieff, J. Forbes, M. C. Burch, W. Rodríguez, and D. P. Dubois (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General 131*(3), 377.

Siewierska, A. (2013). Gender distinctions in independent personal pronouns. ms dryer & m. haspelmath (eds.) the world atlas of language structures online. leipzig: Max planck institute for evolutionary anthropology. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Sinclair, S., C. D. Hardin, and B. S. Lowery (2006). Self-stereotyping in the context of multiple social identities. *Journal of Personality and Social Psychology 90*(4), 529–542.

Slobin, D. I. (1987). Thinking for speaking. In *Annual Meeting of the Berkeley Linguistics Society*, Volume 13, pp. 435–445.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking.". In J. Gumperz and S. Levinson (Eds.), *Rethinking linguistic relativity*, pp. 70–97. Cambridge University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*, pp. 157–191. MIT Press.

Solaz, A. and F. Wolff (2015). Intergenerational transmission of domestic work and gender roles. *Annals of Economics and Statistics 117*(118), 159–184.

Stahlberg, D., F. Braun, L. Irmen, and S. Sczesny (2007). Representation of the sexes in language. In K. Fiedler (Ed.), *Social communication*, pp. 163–187. Hove: Psychology Press.

Starke, K. and A. P. Visser (1994). Sexuality, sexual behaviour and contraception in East Germany. *Patient education and counseling 23*(3), 217–226.

Statistisches Reichsamt (1936). *Statistisches Jahrbuch für das Deutsche Reich*. Online: `http://www.digizeitschriften.de/main=/dms/img/?PPN=PPN514401303_1935` (accessed May 18, 2018).

Stevenson, B. and J. Wolfers (2007). Marriage and divorce: Changes and their driving forces. *The Journal of Economic Perspectives 21*(2), 27.

StGB-DDR (1968). Strafgesetzbuch der Deutschen Demokratischen Republik (criminal code of the german democratic republic).

Sutter, M., S. Angerer, D. Glätzle-Rützler, and P. Lergetporer (2013). The effects of language on children's intertemporal choices. *Firenze: Department of Economics, European University Institute*.

Sutter, M. and D. Glätzle-Rützler (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science 61(10)*, 2339–2354.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics 123*(3), 905–950.

Terwey, M. (2000). ALLBUS: A German general social survey. *Schmollers Jahrbuch 120*(151), 158.

Thaler, R. H. (1988). Anomalies: The ultimatum game. *The Journal of Economic Perspectives 2*(4), 195–206.

Torgler, B. (2003). Does culture matter? Tax morale in an East-West-German comparison. *FinanzArchiv: Public Finance Analysis 59*(4), 504–528.

Torres, A., B. Haas, N. Steiber, and R. Brites (2007). First European quality of life survey: Time use and work-life options over the life course. Technical report, European Foundation for the Improvement of Linving and Working Conditions. Available online: `https://repositorio.iscte-iul.pt/handle/10071/2190` (accessed May 18, 2018).

Treas, J. and E. D. Widmer (2000). Married women's employment over the life course: Attitudes in cross-national perspective. *Social Forces 78*(4), 1409–1436.

Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man, 1871–1971*, pp. 136–179. Aldine.

Trzcinski, E. and E. Holst (2011). A critique and reframing of personality in labour market theory: Locus of control and labour market outcomes. Soeppapers on multidisciplinary panel data research 403-2011, DIW.

UN (2018). Un data: Per capita gdp at current prices – us dollars. `http://data.un.org/Data.aspx?d=SNAAMA&f=grID%3A101%3BcurrID%3AUSD%3BpcFlag%3A1` (accessed May 18, 2018).

UNDP (2015). *Human Development Report 2015. Work for Human Development.*

Van Belle, J. (2016). *Paternity and parental leave policies across the European Union.* RAND. Available online: `https://www.rand.org/pubs/research_reports/RR1666.html` (accessed May 18, 2018).

van der Velde, L., J. Tyrowicz, and J. Siwinska (2015). Language and (the estimates of) the gender wage gap. *Economics Letters 136*, 165–170.

Vesterlund, L. (1997). The effects of risk aversion on job matching: Can differences in risk aversion explain the wage gap? Unpublished manuscript, Iowa State University.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. MASS. Edited by John B. Caroll.

Wickham, H. (2017). tidyverse: Easily install and load the 'tidyverse'. `https://CRAN.R-project.org/package=tidyverse` (accessed May 18, 2018). R package version 1.2.1.

Wieber, A. and E. Holst (2015). Gender identity and womens' supply of labor and non-market work: Panel data evidence for germany. Discussion Paper 9471, DIW Berlin, The German Socio-Economic Panel (SOEP).

Wiswall, M. and B. Zafar (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Witte, J. C. and G. G. Wagner (1995). Effects of employment on East German fertility after unification. In H. Birg and E. Flothmann (Eds.), *Abhandlungen des Demographischen Symposiums des Instituts für Bevölkerungsforschung und Sozialpolitik (IBS-Materialien) Band*, Volume 40, pp. 233–252.

World Bank (2018). The World Bank Data Base: Fertility rates, total (births per woman). `https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?` (accessed May 18, 2018).

Wyrwich, M. (2017). Women and the labour market in East and West Germany: Socialist legacy and pre-socialist tradition. Jena Economic Research Paper 015, Universitaet Jena.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software 16*(9), 1–16.

Zeileis, A. and T. Hothorn (2002). Diagnostic checking in regression relationships. *R News 2*(3), 7–10.

Zhan, C. (2015). Money vs prestige: Cultural attitudes and occupational choices. *Labour Economics 32*, 44–56.

Declarations

Erklärung

Hiermit erkläre ich, Luise Görges, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 22. Mai, 2018

_____

Luise Görges

Eidesstattliche Versicherung

Ich, Luise Görges, versichere an Eides statt, dass ich die Dissertation mit dem Titel: „Economic Analyses of the Impact of Gender Norms on Female Labour Supply“ selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefügten Darlegungen nach §6 Abs. 6 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt: Software Excel, Latex, Stata, R, zTree und die im Quellenverzeichnis angegebene Literatur.

Hamburg, den 22. Mai, 2018

Luise Görges

Selbstdeklaration

Kapitel 1

Kapitel 1 ist in Koautorenschaft mit Miriam Beblo verfasst. Das Kapitel ist im Erscheinen unter dem Titel "On the nature of nurture. The malleability of gender differences in work preferences" in *Journal of Economic Behavior & Organization*. Die folgende Einschätzung in Prozent über die von Luise Görges erbrachte Eigenleistung wurde mit Miriam Beblo einvernehmlich abgestimmt.

- Konzeption: 50%
- Durchführung: 65%
- Manuskripterstellung: 55%

Kapitel 2

Kapitel 2 ist in Koautorenschaft mit Miriam Beblo und Eva Markowsky verfasst. Die folgende Einschätzung in Prozent über die von Luise Görges erbrachte Eigenleistung wurde mit den am Artikel beteiligten Koautoren einvernehmlich abgestimmt.

- Konzeption: 50%
- Durchführung: 30%
- Manuskripterstellung: 40%

Kapitel 3

Die Eigenleistung von Luise Görges für die in Kapitel 3 präsentierte Arbeit liegt bei 100%. Das Kapitel ist veröffentlicht unter dem Titel "The power of love: A subtle driving force for unegalitarian labor division?" in *Review of the Economics of the Household*, 2015, volume 13(1), pages 163-192.

Kapitel 4

Die Eigenleistung von Luise Görges für die in Kapitel 4 präsentierte Arbeit liegt bei 100%.

Kapitel 5

Die Eigenleistung von Luise Görges für die in Kapitel 5 präsentierte Arbeit liegt bei 100%.

Hamburg, den 22. Mai, 2018

_____

Luise Görges