



Gender and the labor market: What have we learned from field and lab experiments? [☆]



Ghazala Azmat ^{*}, Barbara Petrongolo

Queen Mary, University of London, School of Economics Finance, Mile End Road, London E1 4NS, United Kingdom
Centre for Economic Performance (LSE), United Kingdom

ARTICLE INFO

Article history:

Received 11 January 2014
Received in revised form 29 April 2014
Accepted 30 April 2014
Available online 20 June 2014

JEL classification:

J16
J24
J71
C91
C92
C93

Keywords:

Gender
Field experiments
Lab experiments
Discrimination
Gender preferences

ABSTRACT

We discuss the contribution of the experimental literature to the understanding of both traditional and previously unexplored dimensions of gender differences and discuss their bearings on labor market outcomes. Experiments have offered new findings on gender discrimination, and while they have identified a bias against hiring women in some labor market segments, the discrimination detected in field experiments is less pervasive than that implied by the regression approach. Experiments have also offered new insights into gender differences in preferences: women appear to gain less from negotiation, have lower preferences than men for risk and competition, and may be more sensitive to social cues. These gender differences in preferences also have implications in group settings, whereby the gender composition of a group affects team decisions and performance. Most of the evidence on gender traits comes from the lab, and key open questions remain as to the source of gender preferences—nature versus nurture, or their interaction—and their role, if any, in the workplace.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Women have made major inroads in labor markets throughout the past century, resulting in clear convergence in human capital investment and employment prospects and outcomes relative to men (Goldin, 2006). However, while the gender gap in schooling has closed—and even reversed—in most rich countries, there are remaining gender differences in pay and employment levels, as well as in the types of activities that men and women perform in the labor market (OECD, 2002). Women's progress in the labor market has also led to major advances in labor economics, reflecting women's changing role in the economy and identifying the factors behind the remaining disparities with respect to men. The development of novel empirical methods to identify gender differences has been accompanied by broadening perspectives on the gender dimensions of interest.

This paper reviews recent advances in the economics of gender that have been achieved via the experimental approach. We discuss how the experimental literature contributes to a deeper understanding of recurrent questions on gender, as well as to the broadening of research questions towards previously unexplored dimensions of gender differences, and we examine their bearings on labor-market outcomes.¹

The factors driving gender differences in the labor market can be broadly categorized into three forces, which might be interconnected: productivity, preferences and discrimination. By the end of the 1990s, the state-of-the-art work on gender inequalities, summarized in Altonji and Blank's (1999) chapter in the *Handbook of Labor Economics*, had focused mainly on productivity differences related to human-capital accumulation and discrimination as the main sources of gender gaps in wages and hours. At the same time, Altonji and Blank (1999)

[☆] Paper prepared for the EALE 25 Memorial Issue of Labour Economics. We thank Andrea Ichino for very useful comments as a discussant.

^{*} Corresponding author at: Queen Mary, University of London, School of Economics and Finance, Mile End Road, London E1 4NS, United Kingdom.

E-mail addresses: g.azmat@qmul.ac.uk (G. Azmat), b.petrongolo@qmul.ac.uk (B. Petrongolo).

¹ A number of recent papers survey these literatures extensively. Bertrand (2011) discusses new advances in the gender literature, based on both experimental and non-experimental evidence, with an emphasis on gender differences in preferences, gender identity, and women's well-being. Croson and Gneezy (2009) cover lab experiments on gender preferences in detail. The growing usage of field and lab experiments in labor economics is discussed by List and Rasul (2011) and Charness and Kuhn (2011), respectively.

also indicated that a lack of direct evidence on discrimination and gender preferences was a key obstacle to cleanly differentiate among the effects of these three forces. Progress in these areas was hindered by the presence of unmeasured, confounding factors in the study of gender discrimination; and the difficulty of extracting clean information on psychological traits from naturally-occurring data in the study of gender preferences. By providing data explicitly suited to addressing the questions of interest, and allowing tight control on the environment, the experimental approach provides a valuable source of evidence on these and other gender issues.

The experimental approach, both in the field and the lab, represents the most recent addition to the labor economist's toolkit, in the quest for identification of causal effects of interests. As a clear indication of this trend, two chapters in the latest *Handbook* volume are devoted entirely to discussing lessons drawn from experiments in labor economics (List and Rasul, 2011; Charness and Kuhn, 2011). Various factors have contributed to the increased use of experiments in labor research. First, the growing influence of the behavioral literature on economic research has expanded labor economists' perspectives to largely unexplored fields at the border between economics and psychology (see Bertrand, 2011, for a recent survey on new perspectives on gender disparities). Second, labor economists have become more clearly aware of the limitations of observational data in answering old and new questions and have at the same time set higher standards for empirical inference. Finally, economists have become increasingly able and willing to engage in the collection of original data. Given these advances, the experimental approach has made available new data on traditional labor questions and enabled labor economists to address new questions by "letting questions determine the data to be obtained, instead of the data determining the questions that can be asked" (Duflo, 2006).

Experiments in economics have evolved in a number of directions, and, to organize our discussion, it is helpful to classify empirical strategies in labor economics according to the degree of control allowed to the researcher. At one end of the spectrum, there are the traditionally used, naturally-occurring data, over which the researcher has no control—over the information elicited or the economic environment—and for which identifying assumptions are needed to estimate the causal effect of treatment. At the other end of the spectrum, there are laboratory experiments, which use randomization to identify the effect of treatment in the lab (most typically on a subject pool of students) and which allow the researcher to fully control the environment. Thanks to randomization, the causal effect of treatment is identified simply by the difference in mean treatment and control outcomes. Somewhere between these two extremes are (various typologies of) field experiments, which use randomization in a natural-occurring environment—typically on relevant sample pools that may not be aware of their participation in an experiment—and thus allow for a combination of control and realism (Harrison and List, 2004).

In recent decades, the empirical literature on gender has progressed along this conceptual spectrum. Economists have long been interested in the causes and consequences of gender discrimination in the marketplace. Early work on discrimination extensively used the regression approach and decomposition techniques on observational data. However, the increased awareness of this approach's limitations has gradually shifted the emphasis of empirical work on this topic towards field experiments such as audit and correspondence studies, which aim to compare outcomes in the same job for two individuals who are identical in all respects other than gender. While experiments have been used more extensively to study race, rather than gender, discrimination, the experimental approach has in some cases provided clean evidence on gender discrimination in hiring, and represents a promising path for future research. We discuss findings from this approach in Section 2.

More recently, growing emphasis on potential differences in psychological attributes between men and women has shifted the attention of experimental work towards the study of various dimensions of gender preferences, including preferences towards risk, competition,

negotiation and other-regarding preferences (see Croson and Gneezy, 2009, for an exhaustive review of the experimental work on gender preferences). Potential differences in preferences and psychological attributes might offer additional insight into gender gaps in participation to the labor market, in the types of jobs held, and in the performance in a given job. Information on gender preferences is typically elicited in a lab environment, which best isolates one factor of decision, say the attitude towards risk. The recent literature contains numerous examples from this approach, and is discussed in Section 3.

Finally, the emphasis on gender differences at an individual level has led to a recent interest in the role of these differences in collective settings. Higher female representation in high-profile jobs in politics and the corporate sector—partly prompted by regulation such as the introduction of gender quotas in several countries—has led academics and policymakers alike to question the consequences of teams' gender composition for collective decision making. Section 4 links recent empirical evidence from the field and the lab to team work and discusses evidence on the impact of the gender composition of teams on decision making and firm performance.

Section 5 concludes this survey by summarizing the state of the art and discussing current open issues and directions for future research.

2. Discrimination

The study of discrimination, encompassing concepts, measurement and impact, has featured prominently in the gender literature since Becker's (1957) seminal work.² Gender discrimination in the labor market is defined as a situation in which equally productive men and women are rewarded differently, making it necessary to correctly measure differences in productivity in order to pin down the discrimination residual. The early literature has used regression-based methods on observational data—typically labor force or household survey data—to test for discrimination in the labor market. The most common approach (Oaxaca, 1973; Blinder, 1973) consists in decomposing wage (or participation) differentials between men and women into an 'explained' gap, driven by gender differences in observable worker and, sometimes, job characteristics; and an 'unexplained' gap, driven by different returns to given characteristics that are in turn associated with discrimination. Results from this literature are summarized by Altonji and Blank (1999), and point to large unexplained gaps in gender wages and participation rates.

While the existence of an unexplained gap in wages is certainly consistent with discrimination, this measure suffers from two main drawbacks. First, most observational data inevitably lack information on some of the determinants of a worker's productivity, which are nevertheless observed and valued by employers. Thus the unexplained gap is contaminated by unobserved differences in productivity, and whether it provides an upward or downward bias of the true extent of discrimination depends on the sign of differences in such unobservables.³ Second, if pre-labor market investment in human capital is affected by expectations of future discrimination, part of the impact of discrimination is captured by observable productivity differences, and the resulting unexplained gap would underestimate the true extent of discrimination. In the first case, the regression approach would control for 'too little,' while in the second case, it would control for 'too much.'

Experiments are a natural response to some of the weaknesses inherent in the regression approach to discrimination on conventional survey data. As "discrimination is a causal effect defined by a hypothetical *ceteris paribus* conceptual experiment" (Heckman, 1998), the experimental approach allows researchers to approximate the *ceteris paribus* condition by comparing outcomes for otherwise identical men

² See Altonji and Blank (1999, Sections 3 and 4) and references therein.

³ For this purpose, employer–employee matched data provide an improvement over survey data, as they allow researchers to better extract information on productivity of individuals, see Hellerstein and Neumark (2006).

and women on identical jobs, thus removing sources of correlation between gender and other determinants of outcomes of interest. Clearly, this approach addresses the issue of confounding unobservables in survey data estimates, but it would be unable to address pre-labor market discrimination.

The experimental approach in gender discrimination has exploited various forms of audit or correspondence studies to detect discrimination in hiring.⁴ Audit studies on gender compare callback rates and/or offer rates on a given job opening for pairs of applicants—one male and one female—with identical resumes, who are coached to act alike. Neumark (1996) conducts an audit study of hiring discrimination in the US restaurant industry. In his experiment, two male and two female college students are sent to apply for jobs as waiters and waitresses at 65 restaurants in Philadelphia, with a male and female pair applying for the same job at each restaurant. Male and female auditors were made to look identical on paper by creating resumes that were equivalent in terms of educational qualifications and previous work experience, and differences in personality and appearance were minimized. To this purpose, two auditors for each gender were employed in the study, instead of one, to help reduce the impact of traits that are unique to an individual participant, and all participants were instructed to act and dress alike throughout all contacts with the restaurants. Furthermore, while job offer rates are potentially contaminated by differences in personality and appearance, callback rates for job interviews are independent of these, as they are based solely on resumes.

Results show that women are significantly less likely to be invited for interviews and to receive job offers from high-price restaurants. They are also more likely to receive job offers from low-price restaurants, but this effect is imprecisely estimated. While the estimates presented for the high end of the restaurant industry are clearly consistent with discrimination against women, the overall conclusions from this study should be considered with care because of the very small scale of the study and the fact that estimated specifications do not allow for error components that are specific to an applicant/category pair. Thus, if a certain applicant were perceived to be particularly suitable for a high-price restaurant, and this applicant happened to be male, this match-specific component would show up in the estimates as discrimination against women.⁵

While audit studies provide cleaner evidence on discrimination than the typical regression approach on survey data, the attempt to submit pairs of otherwise identical men and women to an experiment may not be fully accomplished. Despite efforts to match auditors on as many characteristics as possible, there may be remaining differences that are valued by employers. More importantly, the fact that auditors are instructed about the purpose of the study may induce conscious or subconscious behavior that is correlated with gender.

Such weaknesses can be overcome in “blind” studies à la Goldin and Rouse (2000). Their well-known study exploits the gradual adoption of blind auditioning by US orchestras during the 1970s and 1980s. This change in hiring practices was prompted by concerns that discrimination would limit the careers of female musicians, especially in the access to great symphony orchestras. In blind auditions, a sound-porous screen is used to hide the identity of the player from the jury, and the comparison of gender differences in hiring rates before and after the introduction of blind auditioning provides some of the cleanest evidence available of discrimination against women. Goldin and Rouse find that blind auditioning increases the probability that a woman is advanced

from the preliminary selection round by 50%, and that it has an even greater effect on the probability that a woman is hired in the final round. Despite the fact that not all their results point in the same direction, the overall conclusion of the paper is that blind auditioning reduced discrimination against women and thus contributed to the recent increase in the presence of women in top orchestras.

Although extremely powerful, blind studies are hard to find, not least because jobs in which prospective employers are able to extract a clear signal of a worker's productivity while their gender is concealed (as it is the case for musicians' positions) are very rare. In the absence of viable blind studies, correspondence studies have recently provided a promising alternative to audit studies. By sending resumes on fictitious applicants instead of real-life auditors, correspondence studies overcome the main weaknesses of the personal approach in audit studies (and are markedly cheaper). Their drawback is that, by construction, they can only identify discrimination in the access to job interviews, and cannot therefore provide any guidance as to discrimination at the job-offer stage. Bertrand and Mullainathan (2004) provide the largest and probably best-known correspondence study of racial discrimination, whereby race is signaled in resumes by white-sounding versus black-sounding names. Although gender differences are not the focus of their paper, Bertrand and Mullainathan detect no evidence of lower callback rates for females than for males.

In the gender literature, the use of correspondence studies has been relatively limited. Riach and Rich (2006) conduct a mid-scale experiment of gender discrimination in hiring in the UK, sending pairs of fictitious applications to job postings in four occupations: accountants, computer programmers, engineers and secretaries. The first two are gender-neutral occupations, while the last two are male-dominated and female-dominated, respectively. The resumes of men and women are carefully matched on qualifications (specific to each occupation). Results show statistically significant discrimination against women in the male-dominated occupations and against men in both the female-dominated occupation and the mixed occupations.

Petit (2007) provides a correspondence study of gender discrimination in the French finance industry. In order to shed light on the role of family constraints in gender discrimination, Petit creates sets of resumes covering various types: young, single and childless; prime-age, single and childless; and prime-age with children. Discrimination against women is detected for young workers in high-skill advertisements, while among prime-age workers, data reveal no evidence of discrimination in either direction.

Audit and correspondence studies are widely viewed as the most compelling avenue to test for discrimination. However, their use for labor market discrimination has been criticized on various grounds (see Heckman and Siegelman, 1993; Heckman, 1998). As noted above, by sending paper applications as opposed to real-life auditors, correspondence studies adequately address concerns of systematic gender differences in unobservables and in auditors' behavior induced by coaching. However, both audit and correspondence studies suffer from a more fundamental criticism, and namely that the assumption of equal group-averages of unobserved productivity is not informative of discrimination when the relevant treatment is not linear in productivity—as is the case in hiring, in which treatment is binary based on a productivity threshold. To give an intuitive example, imagine that men and women apply to a job for which they are well qualified, such that the majority of each group would pass the hiring threshold. In the absence of discrimination, men and women face the same hiring cutoff. Correspondence studies are meant to ensure that, in employers' beliefs, the unobservable productivity component has equal gender mean, and infer discrimination from differences in call-back rates. But suppose for example that, while still imposing equal means, the distribution of unobservables for women is more dispersed than for men. In this case the share of women who pass the hiring cutoff is lower, simply because of the fatter left tail in their unobservables' distribution. The resulting difference in callback rates would be interpreted as evidence

⁴ The discussion that follows refers to field experiments on discrimination. There is also a smaller literature on lab experiments on discrimination (see Anderson et al., 2006, for a survey), which we do not review here, as it has not specifically addressed the gender dimension.

⁵ Results from earlier audit studies (Levinson, 1975; Firth, 1982; Riach and Rich, 1987) are reviewed by Riach and Rich (2002). A general conclusion from this body of work is that women are likely to encounter discrimination in higher-status jobs, but discrimination is not detected in many of the occupations considered. In particular, there is evidence of discrimination against men in female-dominated occupations (Levinson, 1975).

of discrimination against women, but discrimination was excluded by the assumption of equal hiring thresholds.⁶

Work by Neumark (2012) addresses the Heckman and Siegelman (1993) concern by establishing conditions that allow one to recover unbiased estimates of discrimination when group variances differ. His proposed method exploits variation in applicants' observables and requires equal returns to observables across groups. In this case, returns to observables estimated from a correspondence study can be informative of the relative variance of unobservables, which is in turn used to obtain unbiased estimates of discrimination. Variation in observable determinants of productivity is rare to find in correspondence studies, as they typically create one "type" of applicant with certain qualifications and characteristics, and the only randomized characteristic is one that is not intended to have an impact on productivity, such as race or gender. Bertrand and Mullainathan (2004) present an exception to this tendency, by creating high-quality and low-quality resumes. By using their data, Neumark infers higher relative variance of unobservables for black-sounding applicants from the lower returns to a high-quality resume for black-sounding applicants than for white-sounding applicants. When the implied relative variance is used to adjust estimates of callback rates, the estimated effect of race is larger than in the unadjusted estimates, consistent with the low relative hiring threshold assumed in the study. While available experimental estimates of race discrimination seem robust to the main Heckman and Siegelman criticism, it should be stressed that existing audit or correspondence evidence on gender discrimination is scarce, less clear-cut, and still open to this criticism.

Another drawback of both audit and correspondence studies is that they are intended to estimate the average level of discrimination—insofar as they are based on outcomes at randomly selected firms. As pointed out by Heckman (1998), the average level of discrimination is irrelevant for minorities' outcomes in a competitive market, and the only feature that matters is the causal effect of gender at the marginal firm, or at the set of firms where women or ethnic minorities end up applying or working. Discrimination at the average firm, as detected by some audit and correspondence studies, may thus be consistent with no discrimination at the margin. However, while in a frictionless, competitive labor market minority members can escape discrimination even if there is only a single non-discriminating employer, these conclusions would not hold in a market with frictions. With imperfect mobility, minority members are at a disadvantage as soon as any firm is willing to discriminate (see the discussion on this by Manning, 2003, pp. 215–216, and references therein).

While it is not trivial to extrapolate a clear consensus view from the existing body of experimental literature on gender discrimination, it is fair to broadly summarize this literature, concluding that there is evidence of significant discrimination against women in high-status or male-dominated jobs and discrimination against men in female-dominated jobs. Compared to the regression approach, the experimental approach tends to find far more limited evidence of discrimination against women in the marketplace. Different results from the two approaches may be driven by systematic gaps in unobservables in favor of men, which would contaminate the unexplained gap in wages.

Two questions still remain with respect to the existing experimental evidence on discrimination. The first question concerns the nature of the observed extent of discrimination. There exist two main economic models of gender discrimination, one in which employers derive disutility from employing women relative to men ('taste-based' discrimination) and one in which employers use gender to extrapolate a signal of unobserved components of productivity ('statistical' discrimination). Existing evidence on gender discrimination in the workplace is not suited to parsing its nature and typically expresses no claim to it. Progress in

this direction might be made possible by enriching typical field experiment setups with tighter researcher control on elements directly linked to the nature of discrimination. These would include the manipulation of the information available to participants, which would be irrelevant in the case of taste-based discrimination but would instead be highly valued in case of statistical discrimination, and the use of specific games—e.g. the dictator game—that are meant to isolate tastes for discrimination. In this vein, List (2004) shows that the use of complementary framed experiments gives a clear indication of statistical, rather than taste-based, discrimination in the market for sports cards, such that minorities (blacks, the elderly and women) receive lower offers than majorities mostly because dealers use minority membership as a signal for the distribution of reservation values. As framed experiments require tighter control on the environment structure, they tend to depart from real labor market scenarios that are the object of existing audit or correspondence studies, but the design of similar treatments in the workplace could advance the understanding of the nature of discrimination.

The second question regards the mapping between the extent of discrimination revealed by field experiments (especially correspondence studies) and labor market outcomes of interest. Correspondence studies typically give an indication of gender differences in call-back rates in various types of jobs. Their impact on wage and employment differences depends on potential gender outcomes at the job offer stage. In the hypothetical case in which gender differences in call-back rates are completely reversed at the job offer stage, one would observe no gender differences in outcomes such as hiring rates or wages. While this is indeed an extreme scenario, inherent limits of correspondence studies require some care in their use for drawing inferences on employment and wage differentials. In the general case in which differences in call-back rates translate into significant differences in hiring rates, their impact on wage differentials in turn depends in turn on overall worker mobility. If workers receive outside offers very frequently, labor mobility will wash out the impact of differences in hiring rates if there are at least some non-discriminating employers. But as shown by Manning (2003, Chapter 4), substantial search frictions in the labor market translate differences in hiring rates into persistent differences in wages.

3. Individual preferences

The traditional approach in labor economics to understanding gender differences in outcomes has discussed demand-side explanations such as discrimination, and supply-side explanations based on the accumulation of human capital and family constraints. More recently, however, economists have looked into alternative supply-side explanations for gender differences in outcomes that are related to psychological attributes and preferences. This tendency was partly motivated by the need to find alternative explanations for the residual gender gap in outcomes, and eased by the growing influence of the behavioral and psychological literatures in economics.

A great deal of choice exists in factors that shape labor market outcomes. Individual preferences in the spheres of motivation, ambition and competitiveness—to mention a few—drive choice, and, if they are systematically correlated to gender, they contribute to the observed gaps in outcomes. Measuring preferences on observational data is often difficult, if not impossible. Experiments offer a useful methodology that allows one to study behavior and strategic interaction in a controlled environment and, importantly, can be adapted to elicit gender differences in preferences in spheres potentially associated with labor market success.

In this section we discuss experimental findings on gender differences in preferences that may affect job choice and on-the-job outcomes. Gender differences in preferences for risk and competition offer explanations for job sorting—the important first stage of a labor relation. Moreover, differences in bargaining behavior may shed light on

⁶ The opposite result—namely spurious discrimination against men—would be detected if men and women were under-qualified, such that the majority of them would not pass the hiring threshold.

within-job disparities in wages and career progression. More generally, differences in social preferences, potentially featuring at all career stages, could offer interesting insights into how men and women function in the workplace. The quest for evidence on gender differences in preferences, along with the ease of characterizing these preferences in simple lab experiments, has prompted a huge body of experimental work in a relatively short period (see [Croson and Gneezy, 2009](#), for a detailed survey). Below we summarize some of the main findings from this literature, and discuss available evidence on the origin of gender differences in preferences and on their labor market consequences.

Earning differences across occupations and industries have been widely documented in the literature. Occupational and industry segregation of men and women is one of the leading components of gender gaps in earnings ([Altonji and Blank, 1999](#)). As jobs in different sectors offer different arrays of job security, earnings stability and working conditions, systematic gender differences in preferences towards risk and competition have the potential to shape gaps in earnings through job sorting behavior.

Job-related risks (such as job-loss risk or earning volatility) are typically rewarded by higher mean earnings, and high-risk sectors tend to be male-dominated. If women are more risk-averse than men, they end up being overrepresented in jobs with lower-mean and lower-variance salaries. In the lab, simple risk experiments using real and hypothetical gambles have been carried out to understand gender differences in risk preferences. The consistent finding from lab evidence is that men are more risk-prone than women (see surveys by [Eckel and Grossman, 2008](#); [Croson and Gneezy, 2009](#)). While some attribute these differences to the emotional reaction to uncertain situations ([Loewenstein et al., 2001](#)), others report that they are related to confidence ([Soll and Klayman, 2004](#)). Similar conclusions have been drawn from (scant) survey and field evidence ([Dohmen and Falk, 2011](#); [Dwyer et al., 2002](#)).

High-profile careers often develop in highly competitive settings, in which rewards are linked to relative, rather than absolute, performance. Using experiments that ask subjects to sort into environments with varying degrees of competition, we can observe whether there exist gender differences in participation in more competitive environments and, as a second stage, whether there are gender differences in performance in the more competitive environments. In the lab setting of [Niederle and Vesterlund \(2007\)](#), subjects can choose between a piece-rate incentive scheme and a tournament (winner-takes-all) competition. Conditional on performance, women prefer to be compensated piece-rate, while men prefer tournaments. Moreover, by shying away from competition, the well-performing women hurt themselves financially. Other experimental evidence also suggests that men thrive in competitive environments. [Gneezy et al. \(2003\)](#) show that when subjects are compensated by either piece-rate or a winner-takes-all tournament, men perform slightly better than women under piece-rate but substantially better in a tournament setting. Males' performance increases significantly in the competitive setting, while that of the female subjects remains constant. Evidence from a field experiment on school children by [Gneezy and Rustichini \(2004\)](#) broadly confirms lab findings, and also shows that the opponent's gender matters for individual performance. Using short-distance running, they find that girls' speed decreases when they run against other girls, while it slightly increases when they run against boys. Rewards at stake also seem to matter for competitive behavior. [Antonovics et al. \(2009\)](#) find that women perform worse against men than against other women when stakes are low, while they perform better when stakes are high. Finally, lab results show that men and women differ significantly in how they respond to wins and losses from past tournaments, as women's performance is negatively affected by previous losses, while men's performance is only negatively affected when stakes are sufficiently high ([Gill and Prowse, forthcoming](#)).

Another hypothesis for why earnings of men and women differ, even on identical jobs, is that men and women bargain salary differently. In the psychology literature, it has been suggested that women earn

less than their male counterparts because they avoid competitive negotiation and, as [Babcock and Laschever \(2003\)](#) point out, they simply "don't ask" for a pay increase. In a lab setting, the experiment of [Small et al. \(2007\)](#) identifies a clear gender gap in the likelihood of initiating a negotiation, with women asking for a higher payment from the experimenter less often than men. [Bowles et al. \(2007\)](#) study gender differences in negotiation in the presence of an evaluator, who can be male or female. An important result in their study is that women are less likely to start a negotiation in the presence of a male evaluator, and women who initiate negotiations receive systematically poorer evaluations by male evaluators than men.

Ultimatum games are also suited to characterizing attitudes in negotiations, as they involve sharing a given amount between two parties via a single take-it-or-leave-it offer.⁷ [Eckel and Grossman \(2001\)](#) and [Solnick \(2001\)](#) examine gender in ultimatum settings, and both find that men and women make similar offers. However, [Eckel and Grossman \(2001\)](#) find that women are generally less likely to reject offers, while [Solnick \(2001\)](#) concludes that women are more demanding than men when the opposer is female, and less demanding when the opposer is male. Recent work by [Rigdon \(2013\)](#) also finds that women make significantly lower demands than men. However, the gap in demands is mitigated when they are given information on demands made by others in similar negotiations, suggesting that simple pieces of information may shape women's beliefs about the norms of negotiation and reduce earning gaps.

Finally, there has been a great deal of interest in whether men and women exhibit different degrees of social preferences and, in particular, whether other-regarding attitudes such as altruism, fairness or envy may play a stronger role in female than in male decisions. Gender differences in social preferences might help us understand why men and women select into certain sectors or jobs; for example, women tend to be overrepresented in the social sector. Moreover, these differences may have an impact on gender outcomes in a given job, as individuals may take into account their social preferences when negotiating wages and collaborating with coworkers. Social preferences have been investigated in a number of game theoretic settings, including simple ultimatum and dictator games, as well as games that address trust, social dilemmas and public good provision.

In the ultimatum game, the proposer maximizes her (selfish) payoff by offering the smallest amount, and the responder maximizes her own payoff by accepting any offer. When this does not happen, responding subjects have other-regarding preferences, such as altruism, fairness or (negative) reciprocity. No significant gender differences are detected at the offer stage ([Eckel and Grossman, 2001](#); [Solnick, 2001](#)), and only in specific circumstances would women make smaller demands than men ([Solnick, 2001](#); [Rigdon, 2013](#)).

While the ultimatum game setting makes it difficult to establish whether proposers make generous offers because they fear rejection or because they are altruistic, the dictator game overcomes this problem by giving the respondent no choice but to accept the proposer's offer. [Eckel and Grossman \(1998\)](#), [Bolton and Katok \(1995\)](#) and [Andreoni and Vesterlund \(2001\)](#) find that women offer significantly more than men in a simple dictator game, while [Bolton et al. \(1998\)](#) and [Bohnet and Frey \(1999\)](#) find no significant gender differences. Other studies, based on variations of the dictator game, conclude that women are more sensitive than men to equity considerations (see [Croson and Gneezy, 2009](#), Section 3.2).

Sociologists and economists have long believed that differences in trust or "social capital" explain a great deal about the successes and failures of individuals and societies in performance and progress. Insuring against distrust is costly and typically involves formal legal contracts

⁷ In the ultimatum game, one player, the "proposer," makes a take-it-or-leave-it offer such that she divides some amount of money between herself and another person. The second person, the "responder," can accept or reject the division. If the responder rejects, they both get nothing.

or informal pacts through, for example, family solidarity. In experimental economics, simple experiments have been used to elicit trust through the interaction of investors and trustees in trust games.⁸ Several studies find no gender differences in trust (see, among others, Croson and Buchan, 1999), but women's trust is more sensitive than men's to the details of the context. For example, in Eckel and Wilson (2004), some participants are shown a picture of their recipients, while others are given some information; women trust more than men when they have a photo of their counterpart, but less when they have only written information.

Gender differences in social preferences may finally translate into differences in the willingness to contribute to public goods. Eckel and Grossman (2008) survey findings from public good experiments,⁹ and conclude that evidence on gender differences in the provision of public goods is mixed, as some studies find that women contribute more than men, while others find the opposite. Moreover, the comparability of available findings is complicated by variation in the specific experimental designs. In summary, simple dictator or ultimatum games suggest stronger altruistic preferences among women than men, but evidence from other settings is far from conclusive.

While evidence from various experimental settings suggests that women and men may differ in traits that are potentially related to labor market success, the causes—nature or nurture—and the economic consequences of such differences are not entirely understood. Understanding nature and nurture components of gender differences in behavior is a key issue, with clear policy implications. On the nature side, obvious differences between men and women related to childbirth and physical strength, and more subtle differences related to brain structure and sex hormones, may have an impact on how well men and women fare in the labor market. Women's physical traits are likely to have a declining impact on their labor market involvement, thanks to medical progress in contraception, child rearing, or maternal health (Goldin and Katz, 2002; Bailey, 2006; Albanesi and Olivetti, 2009), as well as technological advances biased towards brain relative to brawn skills (Welch, 2000). Nevertheless, a few (non-experimental) studies have detected a significant correlation between exposure to sex hormones and gender behavior (see Dreber and Hoffman, 2007, for an application to risk preferences, and Bertrand, 2011, Section 2.7.2, for a critical review of related work), and others have investigated the relationship between menstruation and absenteeism (Ichino and Moretti, 2009; Herrmann and Rockoff, 2012, 2013).

On the nurture side, preferences may have roots in the education system, the household, or society at large. Recent experimental work has provided evidence of important environmental influences in various contexts. Gneezy et al. (2008) compare attitudes toward competition in patriarchal and matrilineal societies in a lab experiment setting. While men in the Maasai patriarchal society in Tanzania are significantly more likely to compete, the gender differential in competition is reversed in the matrilineal society of the Khasi in India. Symmetric gender gaps suggest that attitudes towards competition may not be driven by biological differences. Using similar lab techniques, Booth and Nolen (2012) study gender differences in risk preferences in single-sex and coeducational schools, and find that girls from single-sex schools have risk preferences similar to boys', while girls in coed schools are more risk averse. In the same context, Booth and Nolen (2009) look at gender differences in the willingness to compete, and find that girls in single-sex schools have competitive preferences similar to boys', while girls in mixed-gender schools choose piece-rate compensation more often than boys.

Household influences may start at a young age through interaction with parents, or later in life through interaction with spouses. Evidence on these mechanisms is scant. Using school-level surveys, Fryer and Levitt (2010) find no evidence of differential parental investment in girls' and boys' math skills, but there is little direct evidence on later influences and in particular on the interplay between the spousal division of labor in the household and women's performance in the labor market. Although not strictly a nurture mechanism, symmetric gender gaps in working hours in the household and the market point to a clear link between women's role as main providers of home production and their labor choices and constraints. Findings from the (non-experimental) labor supply literature show a causal impact of fertility on female labor supply (see e.g. Angrist and Evans, 1998), and more specific evidence on the household division of labor would be welcome to give a richer picture of the labor market impact of women's involvement in the household.

The important next stage is to understand how findings from the lab on gender differences in preferences would map out on labor market outcomes. Typical lab settings such as abstract contexts and relatively low stakes naturally call for real-life validations of lab results. One topic that has received the most attention is that of competitive preferences. Buser et al. (forthcoming) find that the lab measure of competitiveness introduced by Niederle and Vesterlund (2007), based on choice of tournament compensation, explains about a fifth of the gender gap in academic track choices in Dutch secondary schools. As the adopted measure of competitiveness is based, among other factors, on math ability, and the prestige of an academic track is based mainly on math content, part of the observed correlation between competitiveness and choice of prestigious tracks may in principle be driven by variation in math ability. However, the experiment collects independent information on both actual and perceived math ability, and the impact of these on the choice of prestigious tracks is not substantially altered when controlling for competitive preferences. On the other hand, the field experiment by De Paola et al. (2014) on Italian undergraduate students in economics finds no evidence of gender differences in students' willingness to enter a tournament evaluation scheme, or in their performance within the tournament.

Workplace consequences of competitive preferences have been recently addressed by Flory et al. (2010), who explore job application decisions of men and women. In their field experiment, interested job seekers are randomized into compensation regimes with varying incidence of relative performance evaluation. The results show that the proportion of females in the applicants' pool falls with the share of compensation that depends on relative performance. This pattern is not driven by men preferring competitive environments and women not; rather men avoid competitive environments less than women. Important factors are found to explain or attenuate gender differences, including team work, whether the task is female-oriented, and local labor market conditions. While Flory et al. (2010) study gender differences in job entry, Lavy (2013) focuses on differences in performance. In his field setup, male and female teachers participate in a rank tournament based on the test performance of their classes, and his results reveal no gender differences in performance under competition. Finally, using evidence from a UK workplace survey, Manning and Saidi (2010) find the effect of performance pay on earnings is generally modest, and does not differ markedly by gender.

Evidence on other types of preferences, such as bargaining and social preferences, has been mostly descriptive (see, for example, Babcock and Laschever, 2003), with a few exceptions. Leibbrandt and List (forthcoming) use a similar experimental setting as Flory et al. (2010) to randomize interested job seekers into two types of job ads, one in which it is explicitly stated that the salary is negotiable, and one in which no information on wage setting is given. They find that men, unlike women, prefer job environments where no rule of wage setting is stated. In this case, men are also more likely than women to initiate wage negotiations. By contrast, when it is stated that the salary is

⁸ A simple trust game proposed by Berg et al. (1995) involves an "investor" and a "trustee." The investor is endowed with an amount of money, which she can keep or invest, with a given return rate. The trustee is entrusted with the investment and must decide how much of it to keep and how much to return. Trust is the willingness to gamble that another person will reciprocate a risky move, and is measured by the amount invested. Trustworthiness is measured by the amount returned.

⁹ These are essentially n-players prisoner's dilemma games.

negotiable, women are more likely to negotiate. The nonexperimental evidence on this topic typically uses predetermined measures of attitudes to predict earnings gaps. Fortin (2008) uses longitudinal data to investigate the role of greed and altruism, and concludes that differences in these preferences – as recorded prior to labor market entry – are a modest but significant predictor of earnings gaps. Manning and Swaffield (2008) use similar information on a comprehensive set of preferences (risk attitudes, attitudes towards competition, self-esteem and overconfidence, and other-regarding preferences), and show that these can explain around one fifth of the earnings gap. Finally, recent work by Card et al. (2013) shows how gender differences in bargaining power can be identified on matched employer–employee data. They estimate a model of wage determination with additive fixed-effects for workers and gender-specific effects for firms, in which the assumed proportionality between firm effects and rents accruing to each gender is sufficient to identify their relative bargaining power. Their findings imply that women are more likely than men to be employed at low-rent firms, accounting for about 20% of the wage gap, and capture a smaller fraction of firm rents, accounting for 10–15% of the gap.

4. Group preferences and dynamics

A natural progression from the study of individual preferences has been to understand their role in group settings. If different psychological traits lead men and women to make different choices in similar contexts, the gender composition of teams becomes a relevant factor in collective decision-making. Moreover, various studies have shown that individual behavior changes in the presence of people from the same or opposite sex (Gneezy and Rustichini, 2004; Antonovics et al., 2009; Ivanova-Stenzel and Kuebler, 2011), highlighting the role of group interactions.

Economists have shown growing interest in the consequences of gender diversity of teams. Higher female participation to the labor market has implied changing workplace demographics and more gender-diverse teams. In high-profile professions, such as politics or the corporate sector, these changes have been eased by the introduction of explicit gender quotas in a number of countries. With growing representation of women on political committees, company boards and other group settings, understanding the effects of gender diversity is important from an academic, as well as a policy, point of view. Changes in team composition may in general have consequences for corporate and political outcomes. Moreover, higher female representation in leadership roles may accelerate gender equality both in the short run, by attracting high-ability women to such roles, and in the long run, by exposing society to female leadership and accelerating changes in social norms (see for example Beaman et al., 2009).

Despite a large body of lab evidence on individual preferences, the experimental literature on gender and preferences at the team level is relatively scarce. One of the main problems with studying gender and groups is that groups are typically formed in an endogenous way. While experiments can go some way to solve this issue through random assignment into groups, they tend to create an artificial environment, in which it becomes difficult to distinguish group diversity and group dynamics. The ideal setting is one that captures the complexity of endogenous group formation, while allowing the researcher to observe exogenous changes in group composition and identify the causal relationship between group diversity and outcomes.

Recent reforms that mandate certain levels of female representation on boards of directors offer a valuable, quasi-experimental setting to study gender composition of teams and performance. One of the first countries to implement gender quota laws was Norway in 2003. To study the causal impact of female presence on boards on firm performance, Ahern and Dittmar (2012) exploit the Norwegian gender quota reform, which required listed companies to achieve 40 percent female board representation within two years. Approximately 30 percent

of the members of an average board needed to change to be in compliance with the 40-percent quota. However, some firms had a greater proportion of female directors before the quota was imposed and, therefore, faced a smaller constraint than those with fewer women. Given variation in the share of female directors at baseline, the Norwegian setting offers a natural experiment for the study of the impact of gender quotas. Ahern and Dittmar (2012) document important effects of female board representation. In particular, they find that the constraints imposed by the quota implied a decline in stock prices and operating profits.

Understanding the channels through which group composition affects performance is clearly important to design appropriate policy intervention. Ahern and Dittmar (2012) find that the quota setting in Norway led to less experienced boards, as new female directors had less CEO experience and were younger than existing male directors. However, other channels such as gender differences in preferences or group strategic interactions may not be ruled out in their study. Matsa and Miller (2013) study the effects of the Norwegian gender quota on corporate decision-making and conclude that the quota has changed the style of corporate leadership. They compare publicly-listed firms in Norway with a matched sample of unlisted firms in Norway, and listed and unlisted firms in other Nordic countries, and find that most corporate decisions were unaffected after women's board representation increased, but note sizeable differences in firms' employment policies. In particular, firms affected by the quota undertook fewer employee layoffs, causing an increase in relative labor costs. The results suggest that female directors consider labor hoarding a more profitable long-run strategy, or that they have a greater concern for workers' vulnerability to unemployment risk. Finally, Bertrand et al. (2014) study whether gender quotas in company boards had an impact on gender inequalities beyond the board, and find no evidence that the gains in female presence at the top, as induced by mandated quotas in Norway, trickled down to a stronger female presence in lower layers within the companies affected, or outside the companies affected.

Bagues and Esteve-Volart (2010) provide further quasi-experimental evidence on the relevance of gender in teams, by exploiting the random allocation of applicants to selection committees in the Spanish civil service. They analyze the chances of success of male and female candidates to the Spanish judiciary system and find that female candidates are less likely to be hired whenever they are randomly assigned to a committee with a stronger female presence. Further evidence suggests two potential driving mechanisms. First, female evaluators tend to overestimate the quality of male candidates. Second, the presence of women on committees induces male members to increasingly favor male candidates.¹⁰

Recent papers that use field data have also found evidence that the gender composition of a team does influence performance. Apesteguia et al. (2010, 2012) use a large online business game to study how a female presence on a team affects collective choices and performance. The game is played by teams of three, which take the role of a general manager of a beauty-industry company, competing in a market composed of four other simulated companies. The analysis shows that teams composed of three women are significantly outperformed by any other gender combination. Differences in performance are explained by differences in decision-making: all-women teams are less aggressive in their pricing strategies, invest less in R&D, and invest more in social sustainability initiatives, than any other gender combination. Hoogendoorn et al. (2013) use a similar business game setting with Dutch university students, in which students are exogenously allocated to teams, and detect an inverse u-shaped relationship between the share of women and a team's business performance, such that teams with an equal gender mix perform best.

¹⁰ Different conclusions are drawn by De Paola and Scoppa (forthcoming), who find that female candidates for academic jobs in Italy are less likely to be selected when there are no women on the selection committee.

Delfgaauw et al. (2013) provide novel field evidence on gender interactions and firm performance. They run an experiment in a Dutch retail chain to study whether firm performance under competition is affected by the gender of the manager and/or by the gender composition of employees. The experiment introduces short-term sales competition in a random subset of stores, and tournament compensation is offered within sets of treated stores. On average, the introduction of the tournament increases sales growth by five percentage points, and this effect is independent of the manager's gender or the share of female employees. However, the effect of the tournament on sales growth is highest in stores in which the manager and a sufficiently high fraction of employees are of the same gender.

The interaction between gender composition and the altruistic behavior of groups is analyzed in a dictator's game setting by Dufwenberg and Muren (2006). In their game, groups of three people with varying female presence divide a sum of money among themselves and a fourth person. The authors find that female-majority groups give more to the individual recipients and choose the egalitarian division more often than male-majority groups do. They also find that the most generous groups are those with two women and one man, suggesting that the presence of a man in the group triggers an enhanced generosity in the group.

Selection into environments that might be deemed as competitive has been studied at both the individual and group level. Following the work of Niederle and Vesterlund (2007), a number of recent papers have studied entry into incentive schemes in which there exist options for group incentive compensation. Wozniak et al. (2010) find that men are more likely to choose to enter a tournament and that women are more likely than men to choose team pay (though this effect is not statistically significant). Kuhn and Villeval (2013) use a real-effort lab experiment in which participants can choose to work under an individual piece rate or under a group piece rate in which reward depends on the team's performance. They find that women are significantly more likely than men to select team-based compensation and suggest that the same confidence deficit that pushes women out of competition pulls them into teams. Interestingly, however, men become much more likely to join teams—and the gender gap vanishes—when an efficiency advantage to team production is introduced.

While the quota reforms and other field and lab experiments offer valuable insight into the consequences of gender diversity, the literature on this issue is still very limited, not least because it is restricted to a small and select group of women. Gender corporate board quota policies, as well as business games among MBA students, focus attention on women who may not be fully representative of the female workforce. The representation of women in decision-making at lower levels of responsibility can thus help to form a broader picture of the impact of gender diversity and attenuate the stark selection of women at the top.

5. Concluding remarks

Experiments offer a novel and useful methodology that is being used widely in almost all areas of economics. In gender economics, the experimental approach offers a way to answer questions previously believed to be unanswerable because of data limitations, as well as new techniques to cleanly identify mechanisms and results in older topics traditionally studied by labor economists. In this paper we have focused on how experiments help shed light on gender discrimination, as well as on how they have helped our understanding of gender differences in preferences, whether at the individual level or in group settings.

Despite recent advances, several important aspects of gender differences in labor market success have to date not been explored, or have only been partially explored, experimentally, and we believe that there exists a clear scope for further research in several directions.

First, while experimental evidence can identify discrimination in hiring, disentangling different types of discrimination has proved to be challenging. Moreover, to date experiments have offered little insight

into on-the-job discrimination, and on how anticipated discrimination might feed back into individuals' choices.

Second, the psychology and behavioral economics studies on gender preferences and attitudes have helped shed light on gender disparities in choice. While one may conjecture that disparities observed in the lab have implications for labor market outcomes, more direct evidence from the workplace is needed to draw useful conclusions for gender gaps in real markets.

Finally, it is not entirely understood how preferences are formed and, importantly, whether there is scope for policy to change them. For example, gender differences in labor market outcomes are typically mirrored by symmetric gender differences in household roles, leading to the natural question of whether household roles reflect gender preferences and whether, either through nature or nurture, they may shape labor market choices, constraints, and performance. While an established literature, mostly theoretical or non-experimental, studies efficient household bargaining and gender time allocation, further causal evidence is needed to understand the remaining issues on gender inequalities in the household and the labor market.

References

- Ahern, Kenneth R., Dittmar, Amy K., 2012. The changing of the boards: the impact on firm valuation of mandated female board representation. *Q. J. Econ.* 127 (1), 137–197.
- Albanesi, Stefania, Olivetti, Claudia, 2009. Gender roles and medical progress. NBER Working Paper 14873.
- Altonji, Joseph, Blank, Rebecca, 1999. Race and gender in the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 3C. Elsevier, Amsterdam, pp. 3143–3259.
- Anderson, Lisa, Fryer Jr., Roland G., Holt, Charles A., 2006. Discrimination: evidence from psychology and economics. In: Rodgers III, W. (Ed.), *Handbook on the Economics of Discrimination*. Edward Elgar, Northampton, MA, pp. 97–115.
- Andreoni, James, Vesterlund, Lise, 2001. Which is the fair sex? Gender differences in altruism. *Q. J. Econ.* 116 (1), 293–312.
- Angrist, Joshua D., Evans, William N., 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. *Am. Econ. Rev.* 88, 450–477.
- Antonovics, Kate, Arcidiacono, Peter, Walsh, Randall, 2009. The effects of gender interactions in the lab and in the field. *Rev. Econ. Stat.* 91 (1), 152–163.
- Apesteguia, Jose, Azmat, Ghazala, Iriberrri, Nagore, 2010. The impact of gender composition on team performance and decision-making: evidence from the field. UPF Working Paper 1225.
- Apesteguia, Jose, Azmat, Ghazala, Iriberrri, Nagore, 2012. The impact of gender composition on team performance and decision-making: evidence from the field. *Manag. Sci.* 58 (1), 78–93.
- Babcock, Linda, Laschever, Sara, 2003. *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press, Princeton and Oxford.
- Bagues, Manuel, Esteve-Volart, Berta, 2010. Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *Rev. Econ. Stud.* 77, 1301–1328.
- Bailey, Martha J., 2006. More power to the pill: the impact of contraceptive freedom on women's lifecycle labor supply. *Q. J. Econ.* 121, 289–320.
- Beaman, Lori, Chattopadhyay, Raghav, Duflo, Esther, Pande, Rohini, Topalova, Petia, 2009. Powerful women: does exposure reduce bias? *Q. J. Econ.* 124, 1497–1540.
- Becker, Gary, 1957. *The Economics of Discrimination*. University of Chicago Press, Chicago.
- Berg, Joyce, Dickhaut, John W., McCabe, Kevin A., 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142.
- Bertrand, Marianne, 2011. New perspectives on gender. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 4B. Elsevier, Amsterdam, pp. 1545–1592.
- Bertrand, Marianne, Mullainathan, Sendil, 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* 94, 991–1013.
- Bertrand, Marianne, Sandra Black, Sissel Jensen and Adriana Lleras-Munay, 2014. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway". Unpublished.
- Blinder, Alan, 1973. Wage discrimination: reduced form and structural estimates. *J. Hum. Resour.* 8, 436–455.
- Bohnet, Iris, Frey, Bruno S., 1999. The sound of silence in prisoner's dilemma and dictator games. *J. Econ. Behav. Organ.* 38, 43–58.
- Bolton, Gary E., Katok, Elena, 1995. An experimental test for gender differences in beneficent behavior. *Econ. Lett.* 48, 287–292.
- Bolton, Gary E., Katok, Elena, Zwick, Rami, 1998. Dictator game giving: rules of fairness versus acts of kindness. *Int. J. Game Theory* 27, 269–299.
- Booth, Alison L., Nolen, Patrick, 2009. Choosing to compete: how different are girls and boys? CEPR Discussion Paper 7214.
- Booth, Alison L., Nolen, Patrick, 2012. Gender differences in risk behaviour: does nurture matter? *Econ. J.* 122, F1–F114.
- Bowles, Hannah R., Babcock, Linda, Lai, Lei, 2007. Social incentives for sex differences in the propensity to initiate negotiation: sometimes it does hurt to ask. *Organ. Behav. Hum. Decis. Process.* 103, 84–103.

- Buser, Thomas, Niederle, Muriel, Oosterbeek, Hessel, 2014. Gender, competitiveness and career choices. *Q. J. Econ.* (forthcoming).
- Card, David, Cardoso, Ana Rute, Klein, Patrick, 2013. Bargaining and the gender wage gap: a direct assessment. *IZA Discussion Paper No. 7592*.
- Charness, Gary, Kuhn, Peter, 2011. Lab and labor: what can labor economists learn from the lab? In: Ahenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 4A. Elsevier, Amsterdam, pp. 229–330.
- Crosin, Rachel, Buchan, Nancy R., 1999. Gender and culture: international experimental evidence from trust games. *Am. Econ. Rev.* 89, 386–391.
- Crosin, Rachel, Gneezy, Uri, 2009. Gender differences in preferences. *J. Econ. Lit.* 47, 1–27.
- De Paola, Maria, Scoppa, Vincenzo, 2014. Gender discrimination and evaluators' gender: evidence from the Italian academia. *Economica* (forthcoming).
- De Paola, Maria, Gioia, Francesca, Scoppa, Vincenzo, 2014. Are Females scared of competing with males? Results from a field experiment. *IZA Discussion Paper No. 8021*.
- Delfgaauw, Josse, Dur, Robert, Sol, Joeri, Verbeke, Willem, 2013. Tournament incentives in the field: gender differences in the workplace. *J. Labor Econ.* 31, 305–326.
- Dohmen, Thomas, Falk, Armin, 2011. Performance pay and multidimensional sorting: productivity, preferences, and gender. *Am. Econ. Rev.* 101, 556–590.
- Dreber, Anna, Hoffman, Moshe, 2007. "Portfolio selection in utero". Unpublished.
- Duflo, Esther, 2006. "Field experiments in development economics". Unpublished.
- Dufwenberg, Martin, Muren, Astri, 2006. Gender composition in teams. *J. Econ. Behav. Organ.* 61, 50–54.
- Dwyer, Peggy D., Gilkeson, James H., List, John A., 2002. Gender differences in revealed risk taking: evidence from mutual fund investors. *Econ. Lett.* 76, 151–158.
- Eckel, Catherine C., Grossman, Philip J., 1998. Are women less selfish than men? Evidence from dictator experiments. *Econ. J.* 108, 726–735.
- Eckel, Catherine C., Grossman, Philip J., 2001. Chivalry and solidarity in ultimatum games. *Econ. Inq.* 39, 171–188.
- Eckel, Catherine C., Wilson, Rick K., 2004. "Conditional trust: sex, race and facial expressions in a trust game". Unpublished.
- Eckel, Catherine C., Grossman, Philip J., 2008. Differences in the economic decisions of men and women: experimental evidence. In: Plott, C., Smith, V. (Eds.), *Handbook of Experimental Economic Results*. Elsevier, New York, pp. 509–519.
- Firth, Michael, 1982. Sex discrimination in job opportunities for women. *Sex Roles* 8, 891–901.
- Flory, Jeffrey, Leibbrandt, Andreas, List, John, 2010. Do competitive workplaces deter female workers? A large-scale natural field experiment on gender differences on job entry decisions. *NBER Working Paper 15646*.
- Fortin, Nicole, 2008. The gender wage gap among young adults in the United States: the importance of money vs. people. *J. Hum. Resour.* 43, 886–920.
- Fryer, Roland G., Levitt, Steve, 2010. An empirical analysis of the gender gap in mathematics. *Am. Econ. J. Appl. Econ.* 2, 210–240.
- Gill, David, Prowse, Victoria, 2014. Gender differences and dynamics in competition: the role of luck. *Quant. Econ.* (forthcoming).
- Gneezy, Uri, Rustichini, Aldo, 2004. Gender and competition at a young age. *Am. Econ. Rev.* 94, 377–381.
- Gneezy, Uri, Niederle, Muriel, Rustichini, Aldo, 2003. Performance in competitive environments: gender differences. *Q. J. Econ.* 118, 1049–1074.
- Gneezy, Uri, Leonard, Kenneth L., List, John A., 2008. Gender differences in competition: evidence from a matrilineal and a patriarchal society. *Econometrica* 77, 909–931.
- Goldin, Claudia, 2006. The quiet revolution that transformed women's employment, education, and family. *Am. Econ. Rev.* 96, 1–21.
- Goldin, Claudia, Katz, Lawrence F., 2002. The power of the pill: oral contraceptives and women's career and marriage decisions. *J. Polit. Econ.* 110, 730–770.
- Goldin, Claudia, Rouse, Cecilia, 2000. Orchestrating impartiality: the impact of blind auditions on female musicians. *Am. Econ. Rev.* 90, 715–741.
- Harrison, Glenn W., List, John A., 2004. Field experiments. *J. Econ. Perspect.* 42, 1009–1055.
- Heckman, James J., 1998. Detecting discrimination. *J. Econ. Perspect.* 12, 101–116.
- Heckman, James J., Siegelman, Peter, 1993. The urban institute audit studies: their methods and findings. In: Fix, Struyk (Eds.), *Clear and Convincing Evidence: Measurement of Discrimination in America*. The Urban Institute Press, Washington D.C., pp. 187–258.
- Hellerstein, Judith, Neumark, David, 2006. Using matched employer–employee data to study labor market discrimination. In: Rodgers, W. (Ed.), *Handbook on the Economics of Discrimination*. Edgar Publishing, Cheltenham, pp. 29–60.
- Herrmann, Mariësa A., Rockoff, Jonah E., 2012. Does menstruation explain gender gaps in work absenteeism? *J. Hum. Resour.* 47, 493–508.
- Herrmann, Mariësa A., Rockoff, Jonah E., 2013. "Do menstrual problems explain gender gaps in absenteeism and earnings? Evidence from the National Health Interview Survey". Unpublished.
- Hoogendoorn, Sander, Oosterbeek, Hessel, Praag, Van, Mirjam, 2013. The impact of gender diversity on the performance of business teams: evidence from a field experiment. *Manag. Sci.* 59, 1514–1528.
- Ichino, Andrea, Moretti, Enrico, 2009. Biological gender differences, absenteeism and the earnings gap. *Am. Econ. J. Appl. Econ.* 1, 183–218.
- Ivanova-Stenzel, Radosveta, Kuebler, Dorothea, 2011. Gender differences in team work and team competition. *J. Econ. Psychol.* 32, 797–808.
- Kuhn, Peter, Villeval, Marie Claire, 2013. "Are women more attracted to cooperation than men?" Unpublished.
- Lavy, Victor, 2013. Gender differences in market competitiveness in a real workplace: evidence from performance-based pay tournaments among teachers. *Econ. J.* 123, 540–573.
- Leibbrandt, Andrea, List, John, 2014. Do women avoid salary negotiations? Evidence from a large scale natural field experiment. *Manag. Sci.* (forthcoming).
- Levinson, Richard, 1975. Sex discrimination and employment practices: an experiment with unconventional job enquiries. *Soc. Probl.* 22, 533–543.
- List, John A., 2004. The nature and extent of discrimination in the marketplace: evidence from the field. *Q. J. Econ.* 119, 49–89.
- List, John A., Rasul, Imran, 2011. Field experiments in labor economics. In: Ashelfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, vol. 4A. Elsevier, Amsterdam, pp. 104–228.
- Loewenstein, George F., Weber, Elke U., Hsee, Christopher K., Welch, Ned, 2001. Risk as feelings. *Psychol. Bull.* 127, 267–286.
- Manning, Alan, 2003. *Monopsony in Motion*. Princeton University Press, Princeton.
- Manning, Alan, Saidi, Farzad, 2010. Understanding the gender pay gap: what's competition got to do with it? *Ind. Labor Relat. Rev.* 63, 681–698.
- Manning, Alan, Swaffield, Joanna, 2008. The gender gap in early-career wage growth. *Econ. J.* 118, 983–1024.
- Matsa, David A., Miller, Amalia R., 2013. A female style in corporate leadership? Evidence from quotas. *Am. Econ. J. Appl. Econ.* 5, 136–169.
- Neumark, David, 1996. Sex discrimination in restaurant hiring: an audit study. *Q. J. Econ.* 111, 915–941.
- Neumark, David, 2012. Detecting discrimination in audit and correspondence studies. *J. Hum. Resour.* 47, 1128–1157.
- Niederle, Muriel, Vesterlund, Lise, 2007. Do women shy away from competition? Do men compete too much? *Q. J. Econ.* 122, 1067–1101.
- Oaxaca, Roland, 1973. Male–female wage differentials in urban labor markets. *Int. Econ. Rev.* 14, 693–709.
- OECD, 2002. *Women at work: who are they and how are they faring?*. Employment Outlook, Paris.
- Petit, Pascale, 2007. The effects of age and family constraints on gender hiring discrimination: a field experiment in the French financial sector. *Labour Econ.* 14, 371–391.
- Riach, Peter A., Rich, Judith, 1987. Testing for sexual discrimination in the labour market. *Aust. Econ. Pap.* 26, 165–178.
- Riach, Peter A., Rich, Judith, 2002. Field experiments of discrimination in the market place. *Econ. J.* 112, F480–F518.
- Riach, Peter A., Rich, Judith, 2006. An experimental investigation of sexual discrimination in hiring in the english labor market. *BE Press Advances in Economic Analysis & Policy*, 6(2) (Article 1).
- Rigdon, Mary, 2013. "An experimental investigation into gender differences in negotiations". Unpublished.
- Small, Deborah A., Gelfand, Michele, Babcock, Linda, Gettman, Hilary, 2007. Who goes to the bargaining table? The influence of gender and framing on the initiation of negotiation. *J. Pers. Soc. Psychol.* 93, 600–613.
- Soll, Jack B., Klayman, Joshua, 2004. Overconfidence in interval estimates. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 299–314.
- Solnick, Sara J., 2001. Gender differences in the ultimatum game. *Econ. Inq.* 39, 189–200.
- Welch, Finis R., 2000. Growth in women's relative wages and in inequality among men: one phenomenon or two? *Am. Econ. Rev.* 90, 444–449.
- Wozniak, David, Harbaugh, William T., Mayr, Ulrich, 2010. "The menstrual cycle and performance feedback alter gender differences in competitive choices". Unpublished.