In a moment when rising inequalities are fueling populist anger and fears about the future of democratic institutions, the gender angle to the inequality narrative undeniably provides one of the greatest causes for optimism. Across the world, gender gaps in education and labor force participation are on the decline. In the developed world, which will be my focus for much of this lecture, the gender gap in average full-time labor market earnings continues to contract, with women, albeit slowly, improving their representation in the upper echelons of the income distribution.

Despite this positive backdrop, important hurdles remain on the path toward gender equality in the labor market. In particular, I will highlight in this lecture two outstanding, quantitatively significant pain points for women. First, despite the previously mentioned educational gains, women continue to make educational choices that systematically translate into lower expected labor market earnings than men; this is most immediately apparent when it comes to women’s sharp underrepresentation in the math-intensive STEM fields. Second, when they become mothers, women make consequential changes to their labor supply decisions that translate into large and persistent losses in labor market earnings.

Because both of these pain points directly relate to different choices women make in the educational and labor markets, I will then turn to the natural subsequent question: why are women making such different choices? I will argue that this has much to do with sticky stereotypes about gender-specific skills and gender-specific roles. The study of stereotypes in general, and gender stereotypes in particular, has not been central to the economics literature. With a few exceptions I discuss below, the closest economics gets to tackling stereotypes is in statistical discrimination models. The social psychology literature provides a much richer description of what stereotypes are and how they affect us. I will briefly review this literature and explain why it should be more central to our understanding of the remaining pain points on the path toward greater gender equality. In particular, I will argue that a better understanding of how stereotypes operate makes it difficult to conclude that women are making different choices because of different skills or preferences, given that these skills, such as math skills, and preferences, such as preferences for caregiving, are themselves endogenous to the gender stereotypes.

It is useful to take a step back and remind oneself why gender inequalities in the labor market matter. The arguments as I see them are similar to the ones that make us concerned about limited social mobility. There is a fairness and justice argument that tells us that women should share the same set of opportunities as men. There is also an efficiency argument: a few papers (e.g., Cuberes and Teignier 2016, Hsieh et al. 2019) have modeled how greater gender equality in labor market access can improve economic growth. All of these models start from the premise that men and women share at birth similar distributions of labor-market-relevant talent. Starting from that premise, it is obvious that any barrier to women’s entry into the labor market, or into some specific occupations, will be distortive to economic outcomes.

A recent revival of gender-conservative arguments, best exemplified by Jordan Peterson’s wide popularity, is pushing against the central
assumption of these models, that is, the assumption that women and men are drawing from the same distribution at birth. It therefore also seems worthwhile to spend some time in this lecture reviewing what we know about the extent of essential differences in skills and traits between men and women. While there are exceptions, I will argue that a central takeaway from this literature is that average gender differences in skills or traits are typically very small in comparison with the large within-gender variation in these skills or traits.

I will conclude with a discussion of the role researchers, policymakers, and organizations play in reinforcing gender stereotypes and lay out some suggestions for change.

I. Progress

Figure 1 reports on female labor force participation rates around the world in 1990 and 2017, for women between 25 and 54 years of age. At all levels of economic development, female labor force participation rates have been going up. The same is true for the ratio of female to male labor force participation.

Note: The figure reports female labor participation (25 to 54 years old) around the world, as a function of log GDP per capita (purchasing power parity (PPP) based), for the years 1990 and 2017.

Source: World Bank and ILO

These gains in female labor force participation have been accompanied by, and are likely partly a reflection of, rapid improvements in the schooling of girls around the world. The reasons for these improvements in the developing world are multiple, ranging from declines in fertility (and the success of family planning policies) to the elimination of school fees and the use of conditional cash transfers to encourage girls’ schooling (for an overview, see Heath and Jayachandran 2018).

In the developed world, women have not only caught up with men in terms of years of completed schooling but have in fact overtaken them. Figure 2 shows that, by 2017, in all OECD countries except Japan and Germany, a higher share of women than men in the 25- to 34-year-old age bracket had completed at least a bachelor’s degree or equivalent.

Note: The figure reports, across OECD countries and for the year 2017, the share of men and women between 25 and 34 years of age who have completed at least a bachelor’s degree or equivalent.

Source: OECD

---

every dollar earned by the median man.\textsuperscript{2} This average hides substantial heterogeneity, with gender gaps in median full-time earnings above 30 percent in South Korea and below 5 percent in Luxembourg. In all OECD countries except Hungary, Latvia, and Chile, the gender gap in median earnings was lower in 2016 than in 2006, when it averaged closer to 16 percent.

Finally, there is evidence of increased female representation in the upper deciles of the income distribution. Administrative labor income tax data from France (Garbinti, Goupille-Lebret, and Piketty 2018) and the United States (Piketty, Saez, and Zucman 2018) reveal cracks in the glass ceiling. In both countries, women accounted by the mid-2010s for about 30 percent of working-age adults with earnings above the ninetieth percentile of the income distribution, from a base of about 5 percent in the 1970s. Also, in both countries, women accounted by the mid-2010s for about 15 percent of working-age adults with labor earnings above the ninety-ninth percentile of the income distribution.

II. Pain Points

Despite this undeniable progress, why do women’s labor market achievements continue to lag behind those of men? I will focus on two contributing factors. First, while often completing more years of schooling than men, women make educational choices that translate into lower expected labor market earnings. Second, parenthood has sharply asymmetric impacts on labor market outcomes between the genders, depressing mothers’ earnings while leaving fathers’ earnings essentially unchanged.

A. Educational Choices

Figure 3 illustrates one of the key differences in educational choices between men and women. The figure plots, for 2016 and across OECD countries for which the information is available, the share of women versus the share of men who graduated with a tertiary degree in a STEM field. These different educational choices have direct implications for the full-time earnings men and women may expect, as documented in several papers starting with Altonji (1993). For example, using data from the 1993 National Survey of College Graduates, Black et al. (2008) finds that, among women who had labor force attachment similar to men, college major choices could explain more than half of the raw gender gap in log wages.

While there has been some convergence in educational choices over time, the gender

differences that remain continue to have quantitatively large effects on women’s expected earnings compared to men’s. In Bertrand (2018), I use American Community Survey data restricted to individuals who have completed at least a four-year college degree by age 30 and proxy for their earnings potential on the basis of their educational choices. Specifically, for each combination of highest degree and field of study, I compute mean earnings and ninetieth percentile earnings among men who are working full time and have completed that degree–field of study combination. I then compute, by birth cohort, the gender gap in such education-based earnings potential. In the 1950 birth cohort, women chose degrees and fields of study with expected mean (ninetieth percentile) earnings that were about 14 (22) percent below men’s. These gender gaps declined until the cohorts born in the late 1960s but have remained roughly unchanged since. Women born in 1985 (the youngest cohort in the data) chose educational degrees that mapped into about 6 percent lower expected average earnings than men, and 10 percent lower expected ninetieth percentile earnings.

Women’s underrepresentation in STEM is particularly problematic in that STEM occupations are not only high-paying occupations on average but also are occupations where the gender gap in earnings within the occupation is particularly small, as shown in Goldin (2014). Using 2009 to 2011 American Community Survey data, Goldin computes the gender gap in full-time full-year earnings in each of the top 95 or so occupations ranked by male income. She finds that technology and science occupations have the smallest gender pay gaps, while business occupations have the largest. In fact, among individuals who are less than 45 years of age, some of the technology and science occupations are characterized by higher average earnings for women than for men.

B. The Motherhood Penalty

Some of the most compelling evidence of the crucial role children play in explaining the remaining gender gaps in labor market outcomes in developed countries has been produced over the past few years (see, in particular, Angelov, Johansson, and Lindahl 2016; Kleven, Landais, and Søgaard 2019). A summary of this evidence can be found in Kleven et al. (2019), which uses an event study methodology to estimate, in six countries, the impact of the birth of a first child on mothers’ and fathers’ labor market outcomes, which they define as gross labor earnings unconditional of employment status. The analysis reveals remarkable similarities in the qualitative effects of children across these six countries, but also substantial differences in the magnitude of the effects. While first-time parenthood is a nonevent for fathers across the six countries, it is associated with an immediate, sharp, and persistent negative impact on mothers’ average labor market earnings. Kleven et al. (2019) estimates long-run penalties (which the authors define as average annual earning losses five to ten years post–first birth) for first-time mothers ranging from 20 percent in Denmark to 60 percent in Germany; the estimate of this long-run penalty is about 30 percent for the United States (comparable to the United Kingdom).

Other studies, albeit focused on more selected samples, support a causal interpretation of these event studies. For example, Lundborg, Plug, and Rasmussen (2017) uses an instrumental variables strategy based on in vitro fertilization (IVF)-induced fertility variation among childless women in Denmark to estimate the causal effect of having children on career outcomes. The authors show that women who are successfully treated by IVF earn persistently less because of having children.

Additional evidence also drawn from the Danish context shows that the fraction of the overall gender inequality in labor market earnings (unconditional of employment status) that can be attributed to children has increased dramatically over time, from about 40 percent in 1980 to about 80 percent in 2013 (Kleven, Landais, and Søgaard 2019). This increase reflects a combination of two underlying changes: total gender inequality in earnings in Denmark has fallen from 46 percent to 24 percent over that period, and child-related gender inequality in earnings has increased from 18 percent to almost 20 percent. In other words, in Denmark as of the mid-2010s, eliminating the motherhood penalty is broadly synonymous with eliminating average gender differences in labor market earnings.

The mechanisms behind the motherhood penalty are multiple. Many women reduce their labor supply when their first child is born. This includes leaving the workforce altogether
or switching to part-time work. The high cost of childcare in some countries (especially for women with low earnings potential) as well as the difficulty of reconciling full-time work with the length of the school day (typically several hours shorter than the work day) and the school calendar (long summer breaks) surely explains some of these changes in labor supply. Besides the reduction in hours worked, there is also evidence that women experience a decline in hourly earnings subsequent to motherhood. In the search for more flexibility at work (see, for example, Mas and Pallais 2017), many new mothers switch from private sector to public sector employment or turn to self-employment. Goldin (2014) also shows that (even small) gaps in employment and/or reduction in weekly hours worked implies a large drop in hourly wages in high-paying occupations, such as those in the corporate and financial sectors (Bertrand, Goldin, and Katz 2010). Finally, mothers may be more interested in finding work that is closer to their residence and thus have a higher willingness to pay for a shorter commute to work, as suggested in Lundborg, Plug, and Rasmussen (2017) and Le Barbanchon, Rathelot, and Roulet (2019).3

C. Taking Stock

The two pain points I have highlighted above, and have argued to be quantitatively large components of what is left of the gender gap in labor market outcomes between men and women in the developed world, can be traced to women making different choices than men in the education and labor markets. In the extreme, were these two forces the sole drivers of the gender gap in average labor market outcomes, one could rationalize this gender gap despite the absence of any employer-based discrimination against women and with the existence of equal pay for equal (labor market) work.

In fact, the few résumé audit studies that have been conducted, while very limited in the range of jobs and outcomes they consider, generally fail to find that employers directly discriminate against women. Booth and Leigh (2010) studied waitstaff, data-entry, customer service, and sales jobs and found a callback ratio of 1.28 in favor of women. Similarly, in Bertrand and Mullainathan (2004), we found that females in sales jobs receive more callbacks than males, even though the difference is statistically insignificant and economically much smaller than the racial gap in callbacks. Carlsson (2011) sent paired applications for positions in both female-dominated occupations (e.g., teachers and nurses) and male-dominated occupations (e.g., information technology professionals or drivers); women received more callbacks in female-dominated occupations, and there was no significant gender difference in male-dominated occupations.4

If women are lagging behind men not because of labor market discrimination but because of the different choices they make, it is key to understand why they are making such different choices.

One view is that these choices reflect on rational self-interested calculations, with men and women arriving at different outcomes because of essential differences in skills. For example, women may not be as competent at math as men, and this may be why many of them stay away from pursuing STEM careers, or maybe women have a comparative advantage over men in caregiving activities. Women and men may also arrive at these different outcomes because of different preferences. Women, even if they are as competent as men at math, may not like the career opportunities that are open to those with STEM degrees, or they may have a preference for home-based caregiving activities over labor market work. Under this view, there might be no need for any type of corrective action, as there might be nothing to correct.

An alternative view is that the different choices men and women make are instead a reflection of powerful and sticky stereotypes about gender-specific skills and gender-specific roles. In the presence of such stereotypes, men’s and women’s choices might still be driven by self-interested calculations; however, unlike in the first view, the gender stereotypes now shape

---

3Women’s caring responsibilities extend beyond children. While fertility is low (and declining), the need to care for elderly family members is rising as longevity is going up. We do not have as much research assessing how much elderly care accounts for the gender gap in labor market outcomes.

4One audit study (Correll, Benard, and Paik 2007) targeting entry-level and midlevel marketing and business job openings, however, found that mothers get fewer callbacks from employers than childless women do.
the choices that are consistent with self-interest. Establishing which of these two views is most correct, I will argue below, is a daunting task when the differences we can observe between men and women (in skills, traits, or preferences) need not be essential but instead are (at least in part) endogenous to the gender stereotypes.

III. Gender Stereotypes

Economics mainly views gender stereotypes as a manifestation of statistical discrimination (Phelps 1972, Arrow 1973, Aigner and Cain 1977). In statistical discrimination models, the differential treatment of men and women is due to imperfect information, and discrimination is the result of a signal extraction problem. Gender stereotypes in these models are rational beliefs about a group member (a woman or a man) based on the aggregate distribution of a trait or skill in the gender group. In contrast, social psychology views gender stereotypes as special cases of cognitive schemas, or generalizations that individuals make to economize on cognitive resources. Unlike economists, social psychologists are more open to the possibility that these stereotypes need not be accurate or rational.

Another key distinction between economists’ and psychologists’ portrayals of gender stereotypes is with regard to how stereotypes operate. Economics stresses the descriptive nature of stereotypes: gender stereotypes are beliefs about what men and women typically do (or how they typically are), and as such, they provide valuable information to an imperfectly informed decision-maker about what the typical man or typical woman will do (or how they will be).

In contrast, social psychology stresses that gender stereotypes are not only descriptive but also prescriptive. Gender stereotypes are beliefs, shared by men and women, about what men and women should or ought to do (or how they should or ought to be). The prescriptive nature of gender stereotypes motivates men and women to adjust their self-view to what seems appropriate for their gender group. This results in gender identity norms, a concept formally introduced in the economics literature in Akerlof and Kranton (2000).

A common methodology used in social psychology to demonstrate the prescriptive nature of gender stereotypes and, more generally, social stereotypes is to generate exogenous variation in social identity effect by using environmental cues called “primes” that can make a particular social identity more salient, even for a short amount of time. Such priming, typically performed in a laboratory setting, has been shown to cause the views, behaviors, and choices of individuals to change (see Turner 1985; see also Benjamin, Choi, and Strickland 2010 for an application in economics and Oh 2019 for a recent use of this methodology in the field).

The prescriptive nature of gender stereotypes is believed to stem from the high level of contact between men and women (Fiske and Stevens 1993). Relatedly, the prescriptive nature of gender stereotypes also naturally derives from a shared education, with views about gender roles and gender skills passed on from parents to their male and female offspring. Parents, explicitly or implicitly, “gender talk” to their children by emphasizing gender categories and teaching what are appropriate and inappropriate behaviors for boys and girls (Endendijk et al. 2014). A recent paper (Brenøe 2018) shows that this transmission of stereotypes about gender roles appears particularly strong in families with mixed-sex children, suggesting that having an opposite-sex sibling increases exposure to gender-stereotypical socialization in the home. In particular, Brenøe (2018) shows that (first-born) women with a (second-born) brother rather than a (second-born) sister acquire more traditional gender norms, as proxied for by a lower likelihood of completing a STEM degree, a higher likelihood of working in female-dominated occupations, and a higher likelihood of marrying a more traditional partner.

Further supporting the prescriptive nature of stereotypes is a desire to be respected and included as a good group member (Ellemers 2018). Indeed, research shows that women and men who behave in accordance with the stereotype are evaluated more positively than those who behave in a counterstereotypical fashion (Rudman and Phelan 2008, Moss-Racusin, Phelan, and Rudman 2010).

The prescriptive nature of gender stereotypes may help explain why, for example, women who earn more money than their husband increase rather than decrease the amount of time they invest in household work (Bertrand, Kamenica, and Pan 2015). This behavior, while a priori puzzling, can be rationalized once one understands that there is a utility cost associated with going
against the stereotypical expectation of being a good wife. The prescriptive nature of gender stereotypes may also more generally explain why women provide more unpaid care. While they may not, absent the gender identity norms, intrinsically enjoy this unpaid care, or enjoy performing low-promotability tasks (Babcock, Recalde, and Vesterlund 2017), providing this care or engaging in those tasks might still be utility maximizing if it allows the individual to comply with the gender identity norms.

Evidence that women share men’s beliefs about gender roles should come as no surprise to researchers who have studied cross-country differences in gender attitudes or sexism. Figure 4 makes this point. Here, I use data collected for the International Value Survey (IVS) in the 2000s and compute, for each country, the share of men (x-axis) and the share of women (y-axis), 18 years of age or older, who agree with the view that “when jobs are scarce, men have more of a right to a job than women.” While the share of women in a country who agree with the statement is somewhat lower than the share of men, there is high concordance between the genders, even in the most sexist places. Similar patterns emerge when studying the correlation between male and female views on other questions that have been used to measure the strength of sexist attitudes across countries, such as “being a housewife is just as fulfilling as working for pay” or “a working mother can establish just as warm and secure a relationship with her children as a mother who does not work.”

A. Endogenous Preferences

The prescriptive nature of gender stereotypes implies that rational self-interest is itself a function of gender norms. This could be, as stressed above, because the gender identity norms have been fully internalized and are part of one’s self-conception, and hence they directly shape one’s preferences. But this could also be because of concerns about the reputational consequences of deviating from the prescribed behavior. Even if individuals do not fully internalize the gender stereotypes and are not intrinsically motivated to uphold their gender identity, the stereotypes may still affect behaviors and choices because of social image concerns and the reputational costs, material or affective, of deviating from the prescribed behavior for the group.

The fact that women in, say, Egypt broadly share men’s views that women belong at home rather than in the labor market need not imply that they have fully internalized these norms into their preferences. Instead, it may simply reflect that these women know that this is what is expected of them by others. Social image considerations and external pressures to conform may also explain why women ultimately take on more of the unpaid care work; they may not engage in this unpaid work because they are internally motivated to do so but instead because they fear the reputational consequences of not doing so.

A detour, which takes me back to Figure 1, provides a way to illustrate that the possibility of internalization of the gender norms into preferences should be taken seriously. In both the 1990 and 2017 data, there is a U-shaped relationship between economic development and female labor participation. The traditional interpretation of this U-shaped relationship is that it depicts how the process of economic development, and the accompanying structural transformation of the economy, affects women’s participation in the labor market. In the earliest stages of economic development,
economic production is mainly taking place on small farms where the demarcation between home and market production is blurry. As the economy develops, jobs move from the agricultural sector to the manufacturing sector. These jobs are less well suited for women as they happen further away from the home and hence cannot be as easily combined with home production; manufacturing jobs also require more physical strength, which makes them better suited for men. With further economic development comes the rise of service sector jobs; while these jobs might still take place away from the home, they are also physically less demanding and more amenable to women’s participation.

While the dynamics above might have applied in the past, they do not appear to provide a good explanation for the past quarter century. In particular, the data reveal no clear relationship between changes in female labor force participation between 1990 and 2017 and changes in the share of value added that comes from agricultural, manufacturing, or the service sector.5

A better way to make sense of the current U-shaped relationship is to separate countries according to the conservativeness of the views they hold with regard to women’s participation in productive activities outside the home. To do this, I use the IVS data and categorize countries according to whether men, 18 and above, agree with “when jobs are scarce, men have more of a right to a job than women.”6 I separate countries into three groups: low, middle, and high sexism, according to whether, respectively, less than 25 percent, between 25 and 50 percent, or more than 50 percent of respondents agree with that view.

Figure 5 reproduces the 2017 relationship from Figure 1 but separates these three sets of countries. The U shape in Figure 1 appears strongly driven by the presence of a group of mostly middle-income countries with high sexism levels and very low rates of female labor force participation.

One immediately sees from Figure 5 that the set of countries identified as high sexism disproportionately belong to the Middle East and North Africa. In many of these countries,

---

5 See Gaddis and Klasen (2014) for a detailed analysis.
6 As implied by Figure 4, it is irrelevant for this exercise whether one uses men’s answers, women’s answers, or both.
Figure 6 relates evaluative well-being of both men and women to log GDP per capita. As already shown in Deaton (2008), there is a robust positive correlation between economic development and life evaluation in the Gallup data, and the relationship is close to linear. Furthermore, there is no visual evidence that women are particularly lagging behind men in the range of middle-income countries we characterized earlier as more sexist.

This is more formally tested in Figure 7, where I report estimated coefficients on a female dummy interacted with high-, middle-, and low-sexism-level dummies from a regression that includes country fixed effects, year fixed effects, a quadratic in log GDP per capita (PPP based), and the interactions of a female dummy with the three sexism levels. Standard errors are clustered at the country level. Life satisfaction at this time (in about five years) defined as described in notes to Figure 6.

Source: World Bank, International Labour Organization, Gallup’s World Poll

Figure 6. Life Satisfaction and Economic Development, by Gender

Notes: The figure reports average male and female life satisfaction around the world as a function of log GDP per capita (PPP based), averaged across the years 2006 to 2017. Life satisfaction at the individual level is based on the answer to the following question: “Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?”

Source: World Bank, International Labour Organization, Gallup’s World Poll

Figure 7. Gender Gap in Life Satisfaction, by Sexism Level

Notes: Reported in the figure are the estimated coefficients on a female dummy interacted with high-, middle-, and low-sexism-level dummies (as defined in Figure 5), as well as confidence intervals. These coefficients are obtained from the following regression. First, I collapse the Gallup’s World Poll (2006 to 2017) into a gender-country-year panel data-set. I then regress average life satisfaction on country fixed effects, year fixed effects, a quadratic in log GDP per capita (PPP based), and the interactions of a female dummy with the three sexism levels. Standard errors are clustered at the country level. Life satisfaction at this time (in about five years) defined as described in notes to Figure 6.

Source: World Bank, Gallup’s World Poll, IVS

and relative well-being in the developed world despite improvements in education and labor market achievements, are highly consistent with the incorporation of gender stereotypes into preferences. Women in the more sexist countries have internalized that they should not participate in the workforce, and these gender identity norms, while constraining behavior and choice, are not perceived as constraints.

This is also reminiscent of Sen’s (1999) critique of utilitarianism and shows the limits of too much reliance on life satisfaction data as well as the complexities and ambiguities of the concept of preferences that are supposed to motivate people’s behavior (Sunstein 1996). The capability approach, pioneered by Sen and expanded upon by Nussbaum, argues that the ability for policies to improve well-being should be evaluated not just on the basis of “functionings” (achieved states, e.g., material standard of living, having a good job, being healthy) but also on the basis...
of “capabilities” (i.e., opportunities for such achievements). People, Sen explains, should have the freedom to choose from many different combinations of “functionings.” The fact that people do not perceive that they lack that freedom, or the fact that they adapt to misery, does not mean that there is no need for corrective policy, even if revealed preferences do not seem to call for such correction. As powerfully stated in Nussbaum (2001, p. 42), “Even when women appear to be satisfied with such customs, we should probe more deeply. If someone who has no property rights under the law, who has had no formal education, who has no legal right of divorce, who will very likely be beaten if she seeks employment outside the home, says that she endorses traditions of modesty, purity, and self-abnegation, it is not clear that we should consider this the last word on the matter. . . . Women’s development is not clear that we should consider this the last word on the matter. . . . People, Sen explains, should have the freedom to choose from many different combinations of “functionings.”

Figure 8 could be viewed as yet another example of the internalization of gender norms. For this figure, I combine 2015 data from the European Working Conditions Survey (EWCS) and the American Working Conditions Survey (AWCS). Reported in the figure is, by country, the share of female employees who report (to a surveyor) having experienced sexual harassment during the course of work over the past 12 months, as a function of the share of men in that country who agree with “when jobs are scarce, men have more of a right to a job than women.” Contrary to what one might have a priori expected, the relationship is negative: experiences of sexual harassment at work are more common in less sexist countries than they are in more sexist ones. While there might be other explanations, one way to make sense of this pattern in light of our prior discussion is that women’s views as to what is or is not acceptable behavior at work have adapted to the ambient gender norms.

B. Self-Fulfilling Prophecies

Economic models such as Lundberg and Startz (1983) and Coate and Louy (1993) have shown how the descriptive nature of gender stereotypes may make these stereotypes self-fulfilling. Statistical discrimination may depress investment in skills by members of the stereotyped group because members of that group rationally believe that these investments will not be fully rewarded due to statistical discrimination.

However, this mechanism is only an impartial and incomplete view as to how stereotypes can be self-fulfilling. The prescriptive nature of the stereotypes offers another, more mechanical channel. Either because the stereotypes are fully internalized or because of the fear of social sanctions for deviating from the stereotypes, men and women will adapt their behavior to what is expected from their gender group. Importantly, because of the prescriptive nature of stereotypes, stereotypes can be self-fulfilling even if the beliefs of others play no role.

This point has been demonstrated in the stereotype threat literature, with the seminal contribution of Steele and Aronson (1995). This literature has shown that gender stereotypes are associated with superior performance in stereotypical domains and underperformance in counterstereotypical domains (Koenig and Eagly 2005). There is also a large literature documenting the mechanisms,

\[ \text{FIGURE 8. EXPERIENCES OF SEXUAL HARASSMENT AMONG WOMEN, BY COUNTRY-LEVEL SEXISM} \]

Notes: Reported on the y-axis is the share of female employees in each country who report having been subjected to sexual harassment during the course of work over the past 12 months. On the x-axis is the share of men (18 years of age or older) in each country who agree with the statement that “when jobs are scarce, men have more of a right to a job than women.”

Source: EWCS and AWCS 2015 for the y-axis; IVS for the x-axis.

\[ \text{FIGURE 8. EXPERIENCES OF SEXUAL HARASSMENT AMONG WOMEN, BY COUNTRY-LEVEL SEXISM} \]

Notes: Reported on the y-axis is the share of female employees in each country who report having been subjected to sexual harassment during the course of work over the past 12 months. On the x-axis is the share of men (18 years of age or older) in each country who agree with the statement that “when jobs are scarce, men have more of a right to a job than women.”

Source: EWCS and AWCS 2015 for the y-axis; IVS for the x-axis.
often happening without awareness, driving the underperformance in counterstereotypical domains, ranging from inability to fully attend to the task when cognitive resources are spent fighting stress and negative emotions (Schmader and Johns 2003) to being concerned with social image (Stähle et al. 2012). Also, Coffman (2014) shows that, conditional on skill level, women are underconfident and men are overconfident in answering questions in stereotypically male domains. This reveals a catch-22: while counterstereotypical performance might be key to changing stereotypes, the stereotypes themselves make it difficult to achieve this counterstereotypical performance.

One area where we have growing evidence that such self-fulfilling dynamics are at play is with respect to the gender differences in math performance. Women’s underperformance in math is often presented as one of the possible explanations for their underrepresentation in the STEM fields. Data from the Program for International Student Assessment (PISA) show that indeed, in many (but not all) countries, girls score more poorly than boys on standardized math tests. In contrast, girls seem to be about on par with boys in science and perform better than boys on reading tests.

Findings from both micro and macro studies show that much of the gender gap in math performance is socially constructed, rather than a reflection of essential differences in math skills, with documented influences from both the school and the home environments. For example, using Italian data, Carlana (2019) shows that math teachers with stronger implicit stereotypes, as measured by the Gender Science Implicit Association Test, have a negative and quantitatively significant influence on girls’ math performance; furthermore, Carlana shows that teacher stereotypes also have a substantial negative impact on girls’ self-confidence in math. Also, Lavy and Sand (2015) shows that primary school teachers’ biases favoring boys have a positive effect on boys’ achievements and enrollments in advanced-level math courses in high school and a negative effect on these outcomes for girls.

Dossi et al. (2019) finds evidence suggesting that socialization at home is also an important driver of the observed gender disparities in math performance. The authors identify families with a preference for boys by using fertility stopping rules. Girls growing up in a boy-biased family score lower on math tests when compared with girls raised in other families. Also, Nollenberger, Rodríguez-Planas, and Sevilla (2016) examines math scores of second-generation immigrants who may be influenced by the gender attitudes of their parents, which they proxy for using home countries’ gender equality measures; they find that two-thirds of the gender gap in math can be explained by parents’ cultural background.

Various macro studies confirm a strong association between gender attitudes in a country and the size of the gender gap in math performance in that country. For example, Guiso et al. (2008) uses 2003 PISA data and shows a positive correlation between the World Economic Forum’s gender gap index and the gender gap in mathematics. Pope and Sydnor (2010) finds similar patterns across US states. Nosek et al. (2009) collects data on the Gender Science Implicit Association Test from 34 countries and finds that the male advantage in math is higher in countries with higher levels of gendered science-stereotyping. Finally, Lippman and Senik (2018) shows that the gender gap in math performance today is larger in Western European countries than it is in former socialist Eastern European countries; similarly, the gender gap in math today is larger in former West Germany than it is in former East Germany. Lippman and Senik (2018) attributes these findings to the lasting effects of socialist policies to boost female employment, policies that reshaped gender roles and gender identity norms in the old Eastern Block.

Now, to be clear, simply erasing a socially constructed gender gap in math performance is unlikely on its own to erase gender differences in educational choices. In fact, in the United States today, several meta-analyses suggest that the gender gap in math performance is small to nonexistent. One such meta-analysis synthesizes data from state assessments of math skills for more than seven million US students in grades 2 to 11 (Hyde et al. 2008); there was essentially no gender difference at any grade level.

What are the other remaining hurdles? Well, girls may have lower preferences for STEM careers. There is some support for this. For example, Hyde (2014) mentions substantial gender differences in interest in engineering and more modest gender differences when it comes to interest in science and mathematics. However, as extensively discussed above, these gender differences in interests should not be viewed as
essential or hardwired and instead could also be socially constructed. Another often-proposed explanation is that girls may be lacking counterstereotypical STEM role models. For example, Carrell, Page, and West (2010) finds that high-ability women assigned to a female STEM instructor were more likely to pursue a STEM major. In contrast, Griffith (2014) finds that female instructors increased female students’ grades in male-dominated disciplines, but the author did not find that this affected women’s decisions to pursue a major in those disciplines.

Finally, yet another explanation is that women are just good at everything, and happen to be relatively better at reading than math. Using 2012 PISA data on 300,000 students in 64 countries, Breda and Napp (2019) finds evidence that such a comparative advantage might indeed be a key force behind the gender gap in the likelihood to pursue a STEM discipline. In these data, boys outperform girls in math by about a tenth of a standard deviation, and girls outperform boys in reading by about a third of a standard deviation, resulting in a difference between reading and math performance that is about 80 percent of a standard deviation larger for girls than boys. The gender gaps in intentions to pursue math-intensive studies and careers are reduced by around three-quarters when this difference is controlled for. If these results more generally hold true, it might very well be that, once the deleterious impact of gender stereotypes on girls’ math performance is fully redressed, the most effective subsequent policy might be for parents, teachers, and counselors to stop reminding young people too much about what they are particularly good at. Instead, they should devote more of their effort to informing these young people about the longer-term implications, including in terms of income and lifestyle, of different educational tracks.

IV. Gender Differences in Skills and Traits and Accuracy of Beliefs about These Gender Differences

Beyond math skills and interests in STEM, what do we know about the extent of essential differences between men and women in skills, psychological traits, and preferences? And what do we know about the accuracy of beliefs about the extent of these differences? Are gender stereotypes broadly accurate, as typically assumed in economics? Or is there a tendency to exaggerate differences, a view that was once dominant among social psychologists? Despite the clear difficulty, as alluded to above, in measuring the extent of essential differences between men and women on the basis of observed differences, it is nevertheless worthwhile to review what we have learned from the literature that has attempted this exercise. This literature is particularly enlightening when it comes to the second pain point introduced in Section II (i.e., the motherhood penalty).

The standard metric in this literature is Cohen’s $d$-statistic, which is defined as the mean score for males minus the mean score for females, divided by the within-groups standard deviation. A useful way to provide a benchmark for what are large or small values of $d$ is in terms of the percent overlap of the male and female distributions at various values of $d$, assuming that both distributions are normal and have the same variance (Cohen 1988). Under these assumptions, a value of 0.10 for $d$ implies a 92 percent overlap between the male and female distributions and 52 percent of males scoring above the median among females. A value of 0.20 implies an 85.3 percent overlap between the distributions and 54 percent of males scoring above the median among females. This explains why the literature that studies these differences typically interprets values of $d$ at or below 0.20 as consistent with gender similarity, values around 0.50 as moderate, and values above 0.80 as large.

The meta-analyses performed in Hyde (2005, 2014) provide by far the most comprehensive overviews of what we know about the extent of differences between the genders. The main takeaway from these studies is that, just like in the case of gender differences in math performance, but with a few important exceptions, average differences between the genders on a long list of traits are small compared with the within-gender differences. Hyde (2005) provides support for a “gender similarities” hypothesis based on a review of 46 meta-analyses of research on psychological gender differences, yielding 124 effect sizes. The skills and traits covered in this study include cognitive variables, verbal or nonverbal communication, social or personality variables,
psychological well-being, motor behaviors, and constructs such as moral reasoning. Of the 124 effect sizes, 30 percent fall into a close-to-zero range \((d \leq 0.10)\), and an additional 48 percent fall into the small range \((0.11 \leq d \leq 0.35)\).

Hyde concludes her subsequent review (Hyde 2014) of the existing evidence to date with this: “Overall, based on the numerous meta-analyses reported here, there is much evidence in support of the gender similarities hypothesis. Domains in which gender differences are small (around \(d = 0.20\)) or trivial (\(d \leq 0.10\)) include mathematics performance, verbal skills, some personality dimensions such as gregariousness and conscientiousness, reward sensitivity, the temperament dimension of negative affectivity, relational aggression, tentative speech, some aspects of sexuality (e.g., oral sex experience, attitudes about extramarital sex, attitudes about masturbation), leadership effectiveness, self-esteem, and academic self-concept.” However, Hyde also acknowledges exceptions to this general rule: “Exceptions to gender similarities, where differences are moderate \((d = 0.50)\) or large \((d = 0.80)\), include 3D mental rotation, the personality dimension of agreeableness/tender-mindedness, sensation seeking, interests in things versus people, physical aggression, some sexual behaviors (masturbation and pornography use), and attitudes about casual sex.”

The next question is with respect to the accuracy of gender stereotypes. Much of the earlier literature in psychology stressed the cognitive mechanisms via which stereotypes might exaggerate differences between groups. One such mechanism is cognitive biases in information processing, with more attention paid, in the formation of beliefs, to group members who more strongly differentiate the groups (see, for example, Corneille and Judd 1999). Bordalo et al. (2016) provides a formalization of this idea. In particular, Bordalo et al. (2016) micro-founds the idea that stereotypes can exaggerate true differences between groups by modeling stereotype formation as a consequence of Kahneman and Tversky’s (1972) representativeness heuristic. In that model, individuals, when forming beliefs about a group, recall from memory only the types that are most representative of that group relative to an alternative group, instead of recalling the full distributions. The authors define the types that are most representative of a group relative to another group as the types whose relative frequency is much higher in the group than in the alternative group. Hence, by construction, stereotypes exaggerate the distinctive traits of the group they represent: true but small differences between the groups (a kernel of truth) are amplified by this belief-formation process.

The research that has tested the accuracy of gender stereotypes across a variety of skills and traits has yielded mixed results. Some of the findings are consistent with accurate beliefs, while others show overestimation and yet others show underestimation of true group differences (Jussim, Crawford, and Rubinstein 2015).

A recent paper (Eyal and Epley 2017) provides one way to make sense of these seemingly inconclusive results, which can be tied back to the cognitive mechanism described above. The specific hypothesis that Eyal and Epley propose to test is that stereotypes will be more likely to exaggerate true gender differences for skills or traits that are presumed to most differentiate the genders; in contrast, stereotypes will be more accurate about gender differences for skills or traits that do not immediately come to mind when thinking about what differentiates a man from a woman. The past literature might have been inconclusive because it measures the accuracy of stereotypes across a wide range of skills and traits, some of them top of mind when thinking about what differentiates a woman from a man but others not.

What is top of mind when thinking about what differentiates a woman from a man? Eyal and Epley consider attributes related to social sensitivity, self-esteem, and happiness. Their survey results show that women are presumed to differ more from men especially with respect to attributes related to social sensitivity (e.g., empathy, compassion) and also somewhat with respect to self-esteem; in contrast, happiness or depression are not top of mind when thinking about what is a male trait or a female trait.

Eyal and Epley then test the accuracy of subjects’ beliefs about gender differences in social sensitivity, self-esteem, and happiness. Subjects (males and females) completed social sensitivity, self-esteem, and happiness tests, which the

---

8 Two recent papers in economics have empirically tested, and rejected, the accuracy of gender stereotypes: Sarsons (2017) and Bohren, Imas, and Rosenberg (2019).
authors use to measure gender differences in these traits. Subjects were also asked to predict what the measured gender differences on these tests would be. Consistent with the original hypothesis, the subjects largely overestimated gender differences in social sensitivity but did not overestimate gender differences in happiness. Importantly, and consistent with the results in Hyde (2014), there was a significant difference between how women and men did on the social sensitivity tests, with women scoring higher than men; that is, there was a kernel of truth underlying the beliefs. Indeed, social sensitivity is frequently described as the “essential difference” between men and women (Baron-Cohen 2003). However, despite this kernel of truth, or rather because of this kernel of truth, beliefs about the magnitude of this difference were exaggerated.

These findings, taken at face value, seem particularly relevant to our understanding of the motherhood penalty. There is a kernel of truth to the fact that women might indeed be better at taking care of children than men,justifying why more mothers than fathers would be putting a brake on their labor market activities to devote more time to caregiving in the home. But the expectations others place on women, and women place on themselves, as to how much empathy, compassion, and warmth they should show are out of sync with the average woman’s biological disposition to such social sensitivity. In other words, women care more and better, but not as much as men and women think they do or believe they should.

There is recent complementary evidence that stereotypes related to women’s social sensitivity skills have been particularly sticky over time. Eagly et al. (forthcoming) conducts a meta-analysis of nationally representative US public opinion polls on gender stereotypes spanning the period from 1946 to 2018. The polls allowed the authors to assess, over time, whether the following traits are perceived to be more true of women than men, more true of men than women, or equally true of both: communion (affectionate, emotional), agency (ambitious, courageous), and competence (intelligent, creative). The communion trait roughly maps into the social sensitivity traits above. The findings show that while men started with a competency advantage, this has disappeared over time, with growing shares of survey respondents either believing that competency is a female trait or believing in gender equality in competency. In contrast, communion was always perceived as a female trait, and the belief that it is more of a female trait than a male trait has in fact only increased over time.

One might attribute the stickiness of the stereotypes regarding women’s superior social sensitivity to women’s continued overrepresentation in caregiving activities (both at home and via the remaining occupational segregation in the workplace); in contrast, women’s educational progress and increased labor force attachment may explain the positive update over time in the beliefs about their competence. However, another interesting conjecture for the dynamics observed in these polls is based on the observation that the remaining stereotypes about women are mainly positive ones. Positive stereotypes (women are nice, women are caring) might be much harder to undo than negative stereotypes (women are incompetent) because they are less objectionable. They may receive more social support (or at least less opprobrium) because the harmful consequences of these stereotypes are less obvious to grasp.

V. How to Combat Remaining Gender Stereotypes

A. The Role of Research, Media, and Education

While the discussion above suggests support for a gender similarities hypothesis across a wide range of domains, with within-gender variation swamping average differences between men and women, this fact would likely come as a surprise to many. Tinsley and Ely make this point quite powerfully in a recent piece (Tinsley and Ely 2018). They note that despite widely held beliefs among many economists that women are less confident than men, have lower preferences for risk, and/or are less willing to negotiate, traits that have been of particular interest to economists, meta-analyses reveal that the gender gaps in these three traits are minimal and, when they arise, are highly context dependent. Academic journals are more interested in publishing research that finds significant differences between the genders than research that finds no such differences. A focus on statistically significant differences in means also often fails to quantify how big these differences are when
compared with the underlying individual-level variation. Researchers, in economics in particular, face limited incentives to undertake replication work, and meta-analyses are often relegated to lower-tier journals, if they are performed at all. The preregistration movement is only now slowly expanding to laboratory-based studies in economics and psychology. All of this translates into more spotlight on research that emphasizes gender differences, with the associated implications of making differences between the genders appear bigger than they truly are, ultimately reinforcing gender-stereotypical thinking. It is crucial to redesign the research production function and researchers’ incentives so that a finding of no gender difference is considered as valuable as a finding of a gender difference.

The media further amplifies publication biases because it understands that its audience will pay more attention to findings of gender differences than it will to nonfindings. There is also suggestive evidence that television programing, online content, and search algorithms contribute to further reinforcing stereotypical expectations. For example, Kay, Matuszek, and Munson (2015) compares the percentages of women who appear among the top 100 Google image search results for a set of occupations to Bureau of Labor Statistics data on the share of women in these occupations and finds large discrepancies in some occupations (but also pretty good calibration in others). For example, 25 percent of the people depicted under the “author” search term were women, while women in fact account for 56 percent of US authors; 64 percent of the people depicted under the “telemarketer” search term were women, even though this occupation is pretty evenly split between men and women.

In addition, the advertising industry appears to cater to the ambient culture regarding gender roles, likely because it is easier (and hence cheaper) to make ads that rely on old clichés. A study that coded how gender roles are portrayed in advertisements in the three largest newspapers of both Italy and the Netherlands (Tartaglia and Rollero 2015) finds that female characters in these ads are more frequently depicted in “decorative” roles; they are also more frequently sexualized, especially in the more gender-unequal Italian context. A few European countries, including Belgium, France, Finland, Norway, and, more recently, the United Kingdom, have introduced regulations aimed at limiting the use of gender stereotypes in advertising. The Advertising Standard Authority, the UK regulator, ruled in 2018 that “advertisements must not include gender stereotypes that are likely to cause harm, or serious or widespread offense.” This includes ads that show men or women “failing to achieve a task specifically because of their gender,” such as a man’s failure to change diapers or care for children.

Finally, the materials that parents and educators rely on to raise their children also often contribute to reinforcing stereotypes. This includes toys and books. For example, there is evidence that toys are more divided by gender today than they were in the past, with toys marketed to boys providing greater opportunities to practice mathematical reasoning and spatial cognition skills (see, for example, Terlecki and Newcombe 2005). Various reports (see, for example, Blumberg 2015) document the under-representation of girls in textbooks throughout the world and, conditional on representation, the overwhelming tendency for them to be depicted in stereotypical roles and as passive actors, with only slow improvements over time. The same holds true when it comes to the case study materials used to educate business leaders in the United States (see Symons and Ibarra 2014, Soule, Drabkin, and Mackenzie 2019).

B. The Role of Family and Tax Policies

Direct exposure to the proscribed, counterstereotypical behavior is key to changing gender stereotypes across generations. For example, Fernández, Fogli, and Olivetti (2004) shows that World War II was a key force in reshaping gender stereotypes in the United States as it exposed more children to having a working mother. In particular, the authors show that the wives of men whose mothers worked because of wartime mobilization were themselves significantly more likely to work. Others have documented transmission from mothers to daughters (Fernández 2011), via local exposure to employed women (Fogli and Veldkamp 2011), and over multiple

---


10 But see also Rose (2018).
generations (Gay 2019). Bastian (forthcoming) shows that the introduction of the Earned Income Tax Credit in 1975 incentivized mothers to enter the workforce and that states with larger responses also experienced greater swings in beliefs toward gender equality after 1975.

Given this, it seems obvious that policies that take women out of work indirectly reinforce gender stereotypes and thus should be avoided. Olivetti and Petrongolo (2017) provides a thorough summary of what we know (and don’t know) about the design of family policy. When it comes to the length of maternity leave, the cross-country evidence strongly suggests that too much generosity might be detrimental to mothers’ labor market success, by hindering mothers from reentering employment on the same track as they were on prematernity. In an analysis that extends Ruhm (1998), Petrongolo and Olivetti (2017) measures the impact of family policies in a panel of 30 developed countries over the period from 1970 to 2014. The authors show that while moderate job-protected leave entitlements (up to one year) might be beneficial to the employment of low-skilled women, more generous leave entitlements hurt the employment of both high- and low-skilled women. Among the higher skilled, more generous maternity leave entitlements are also associated with larger gender gaps in earnings.

Besides family policy, features of tax policy may also disincetivize women’s labor supply. While there has been quite a lot of work on the impact of tax progressivity on married women’s labor force participation (see, for example, Eissa 1995), other aspects of the tax code had been until recently underexplored. A series of recent papers (Bick and Fuchs-Schündeln 2017, 2018) shows that elements of joint income taxation of couples in the tax system might be particularly detrimental to married women’s work. The combination of joint taxation and progressivity implies a marginal tax rate on the primary earner (typically the husband) that is often much lower than the marginal tax rate on the secondary earner (typically the wife). This is in contrast with a separate taxation system where each spouse’s marginal tax rate increases only in his or her own income. As an illustrative example, Bick and Fuchs-Schündeln (2018) shows that the marginal tax rate a typical married women faces in the United States can be as high as in Sweden despite the higher average tax rates in Sweden than in the United States because the US tax code has some elements of joint taxation while Sweden has adopted independent taxation. The marginal tax rate faced by married women in Germany is particularly high because Germany has both high average tax rates and joint taxation of couples: the average income tax rate for single women in Germany is about 30 percent, compared with about 50 percent for married women.

Bick and Fuchs-Schündeln (2017) uses a macro model to quantify how much hours worked by married women would change if all countries adopted separate taxation, holding constant the average tax burden of married couples. The authors estimate large increases in married women’s labor supply associated with a switch out of joint taxation in places like the United States (about an 8 percent increase in hours worked) and Germany (about a 25 percent increase). Consistently, Selin (2014) shows that employment grew considerably among women married to high-income earners in Sweden after the country replaced its joint income taxation system (with steep progressivity) with separate taxation in the early 1970s. These disincentive effects built into the tax system might be especially large for married women with children, as their labor supply might be particularly elastic.

Instead of family and tax policies that discourage women’s, and especially mothers’, work and reinforce stereotypical thinking, one should prefer more gender-neutral policies that make it easier for women to combine labor market work and family responsibilities. In the realm of family policies, the macro evidence summarized in Petrongolo and Olivetti (2017) strongly suggests that one policy that is associated with more equal gender outcomes (employment and wages) across all skill groups is higher government spending on childcare and early education.

Finally, policies that directly try to counter stereotypes should be encouraged. Family policy could be used to encourage more fathers to take on caregiving roles. A few OECD countries have been going down this path. Sweden was the first country to introduce explicit paternity leave rights in 1974, allowing mother and father to share six months of parental leave. Since then, several nations have enacted policies that explicitly incentivize fathers to take up some of the parental leave. These nations have instituted “daddy quotas” that reserve some paid
parental leave for fathers. These “daddy quotas” are nontransferable: if the leave is not taken up by the partner of the main leave-taker, it is lost. Studies from Sweden, Norway, and Quebec, where such “daddy quotas” have been enacted, show that most fathers tend to take up their quota (never much more than a couple of months) but rarely more than their quota (Dahl, Løken, and Mogstad 2014; Ekberg, Eriksson, and Friebel 2013; Patnaik 2019).

However, the financial incentives built into these policies may not always be enough to undo the influence of beliefs about gender roles. For example, Japan and South Korea provide the longest paid father-specific leaves in the OECD (about 12 months). Yet, take-up among Japanese and Korean fathers is very low. According to official statistics from Japan’s Ministry of Health, Labour, and Welfare, only about 6 percent of working fathers applied for or took paternity leave in 2018 despite very generous replacement rates, and a majority of those who took a leave did so for less than a week.11

While these figures give the impression of an uphill battle when it comes to using policy to create counterstereotypical behavior in places where these stereotypes might be particularly strong, some recent research provides a more optimistic outlook. This research hints at the possibility that fathers in Japan may not take up their reserved leave not so much because they feel that it would conflict with their internal self-views but instead because of image concerns; furthermore, and most importantly, these image concerns might be exaggerated. In particular, Miyajima and Yamaguchi (2017) shows that many men in Japan overestimate male peers’ negative attitudes toward paternity leave; among these men, many state that they would like to take paternity leave but are unwilling to because they believe that other men would judge them negatively. Hence, the low rate of paternity leave-taking in Japan might be in part a reflection of pluralistic ignorance, that is, a situation where individuals personally reject a norm but incorrectly believe that many other people accept that norm and hence end up following that norm because of the fear of social sanctions. Recent research in Bursztyn, González, and Yanagizawa-Drott (2018) shows that pluralistic ignorance might also be holding back Saudi men’s willingness to let their wives participate in the labor force. Such findings, if they generalize, indicate that information policy could be a powerful complement to family policy when it comes to speeding up the production of counterstereotypical behaviors in the child-rearing domain.

C. Organizational Practices

Just like some aspects of family and tax policies may have the counterproductive effect of reinforcing gender stereotypes, and in particular stereotypes regarding women’s greater interest in the home than in the workplace, organizational practices may also have such an effect. In particular, changes in the structure of work may be making it more difficult for women, and particularly mothers, to fully realize their labor market potential. This is true among both the more and less skilled segments of the labor market. For example, studying data from the 1980s to the late 2000s, Cha and Weeden (2014) shows increasing shares of Americans working 50 hours or more per week, as well as increasing hourly wage returns to such “overwork.” These trends are particularly strong in professional and managerial occupations. Similarly, while Goldin (2014) establishes that the elasticity of earnings to hours worked is particularly large among higher-paying skilled occupations, I further show that there has been a trend up in this elasticity over time, suggesting increasingly inflexible workplaces for the college educated (Bertrand 2018). Technological change might have been an important force here, with more opportunities to work out of the office and across time zones.

The workplace is also becoming increasingly unaccommodating to less skilled female workers with responsibilities at home. In the United States at least, there has been a rise in irregular, unstable, and unpredictable work hours in lower-paying occupations (Lambert 2008, Boushey and Ansel 2016). Technological change, such as advances in just-in-time scheduling software, may have played an important role in driving these trends, enabling employers

---

to reduce labor costs with finer tuning of their hour-by-hour workforce to expected needs.

There might be additional harmful effects associated with the rise of irregular work. These workplace practices might weaken the bargaining power of women who depend on a more predictable and stable schedule and make these women particularly vulnerable to abusive behavior at work. Figure 9 confirms this possibility. Here, I use the micro data from the EWCS (2010 and 2015) and the AWCS (2015), restricted to the sample of female employees. I regress the likelihood of having experienced sexual harassment during the course of work over the past 12 months on each indicator. The regressions also control for country fixed effects, year fixed effects, a quadratic in age, dummies for the usual number of days at work, and education dummies.

These changes in work practices might be symptomatic of a collective action problem that public policy may need to address. Each organization is doing what is best for itself to minimize its costs, without internalizing the implications of its practices for the overall size, quality, and experience of the workforce. From the narrower perspective of this lecture, these changes are only making it harder for women to defeat the remaining stereotypes regarding gender roles.

Finally, organizations have a role to play in making sure that gender stereotypes do not interfere with their human resource decision-making process. While many organizations have been promoting a “diversity and inclusion” agenda, inclusive practices should be the main driver of this agenda. This is partly based on the recognition that stereotypical thinking can handicap both women and men in the labor market. For example, men are underrepresented in the growing human, health, and social services sector of the economy. These occupations emphasize communality, care, and social skills, and it is possible that gender stereotypes are standing in the way of greater male inclusion in these roles (Croft, Schmader, and Block 2015). This is particularly important given that the labor market is increasingly rewarding these social skills (Deming 2017) and that these occupations, because they are better protected from technological change, are likely to represent a growing share of the economy in the future (Webb 2020).

But maybe more importantly, too much focus on diversity objectives for diversity’s sake only reinforces the tendency to look at job candidates or employees via the lens of group membership (male versus female), while instead, organizations should try to abandon this lens and focus on the individual, whatever his or her group membership.

Besides providing education about the nature of stereotypes, including their descriptive and prescriptive components, organizations should promote strategies that push decision-makers to move beyond the social categorization of job
candidates and employees, such as individuation training or perspective-taking training (for an overview, see Bertrand and Duflo 2017). Organizations should also make the best use of the growing set of sociotechnological solutions that exist to combat humans’ tendency to implicitly rely on stereotypes to make decisions, especially when they are rushed, tired, stressed, distracted, and more generally lacking the cognitive resources to decategorize and “individuate” (see, for example, Bertrand, Chugh, and Mullainathan 2005). There are more and more tools organizations can rely on, often leveraging the insights from behavioral science (Bohnet 2016), to limit the extent to which gender bias creeps into the employee recruitment and evaluation process, from writing a job description to making promotion decisions. For example, a few recent studies have shown how responsive the composition of the pool of job applicants can be to simple changes in the wording of job vacancies (Del Carpio and Guadalupe 2018, Abraham and Stein 2019), confirming the value of the recent products that have been developed to help companies write more gender-neutral job postings.12 While still a work in progress, as exemplified by the recent debacle surrounding the Amazon artificial intelligence recruiting tool,13 there is much to look forward to when it comes to the use of algorithms and machine-learning tools to support organizations in their human resource decisions.

VI. Concluding Remarks

While my focus throughout most of this lecture has been on the remaining pain points on the path toward gender equality, it is important to remember, as I conclude, that the starting message was one of steady progress toward that goal. At the same time, there are some worrying signs of renewed gender conservatism that some have attributed to men’s declining economic and social status, which may have been amplified by a financial crisis that hit male jobs hardest (see, for example, Cassino 2018). It may not necessarily be smooth sailing ahead, especially if economies struggle to continue creating good quality jobs.

Another message I hope this lecture delivered is that we can be better as a discipline if we continue to enrich our models with insights from other disciplines that are concerned with many of the questions that motivate us. Imports from cognitive psychology have been transformational to the field of economics and have improved our ability to understand and predict human decisions. We should not stop there. For example, economics also has a lot to gain from more systematically enriching its models with insights from social psychology.

Finally, it is tempting to say, and many have, that the world would be a kinder—and, in the long-run, better—place if more women were in charge, with less hate, less greed, and/or more sustainable policies. While I find it hard not to sympathize with this argument, I also fear that it is ultimately counterproductive, as it takes us back to the same stereotypical thinking trap. Instead, we should strive for a better allocation of talent in society, whether that talent is male or female.

REFERENCES


Carrell, Scott E., Marianne E. Page, and James E. West. 2010. “Sex and Science: How Professor Gender Perpetuates the Gender Gap.”


Symons, Lesley, and Herminia Ibarra. 2014. “What the Scarcity of Women in Business


