Competition and Gender Inequality: A Comprehensive Analysis of Effects and Mechanisms

BSE Working Paper 1292| September 2021

Klarita Gërxhani, Jordi Brandts, Arthur Schram

bse.eu/research

Competition and Gender Inequality:

A comprehensive analysis of effects and mechanisms*

Klarita Gërxhani, Jordi Brandts, Arthur Schram

September 13, 2021

Abstract

This study provides a comprehensive analysis of gender differences in performance caused by two different dimensions of competition –rivalry for resources and status ranking. It also examines two mechanisms behind such differences: (1) gendered beliefs about performance differences in competitiveness; and (2) prescriptive stereotypes about women having to show warmth towards others. The results indicate that in the absence of any competitive dimension men and women perform equally well. Any competitive dimension, however, leads to women doing worse than men. These results are explained by men's beliefs that they are better than women, and by women's adherence to a prescribed stereotype of not harming others. Gender differences under competition seem to be endogenous to situational contexts, just like they are without competition.

Keywords: gender inequality, competition, status characteristics theory, mechanisms, experiments

JEL codes: C91, J16

***Acknowledgments**

We thank the EUI (Grant IP53), the Spanish Ministry of Economics and Competitiveness through Grant: ECO2017-88130 and through the Severo Ochoa Program for Centers of Excellence in R&D (CEX2019-000915-S), the *Generalitat de Catalunya* (Grant: 2017 SGR 1136), and the Research Priority Area Behavioral Economics of the University of Amsterdam for financial support. We thank Arnout van der Rijt, Robb Willer, and Ezra Zuckerman, for helpful suggestions at various stages of this project.

Authors

Klarita Gërxhani (corresponding author)	Jordi Brandts	Arthur Schram
Department of Political and Social Sciences European University Institute Via dei Roccettini 9 50014 San Domenico di Fiesole (FI), Italy phone +39-055-468.5470 Klarita.Gerxhani@eui.eu	Institut d'Anàlisi Econòmica (CSIC) and Barcelona GSE Campus UAB 08193 Bellaterra (Barcelona) Spain phone +34-93-580.6612 Jordi.Brandts@iae.csic.es	Department of Economics, EUI, Italy *and* CREED, Amsterdam School of Economics, University of Amsterdam, P.O. Box 15867 1001 NU Amsterdam The Netherlands phone +31-20-525.4252 Schram@uva.nl

INTRODUCTION

Competition is omnipresent and almost unavoidable in both professional and recreational life. We regularly compete for jobs, mates, leadership, wealth, recognition, and more. Having to compete may, however, affect how people perform. While some excel under competition, others underperform and fall behind. If there are systematic differences in how men and women respond to competition, then an increasing competitive environment could have important consequences for gender inequality. This is the main motivation underlying our study. Consider the labor market. Though gender equality is arguably promoted by an increased labor participation by women (e.g., Hobson 1990), this increase also means that women are more often than before entering competitive environments. Competition may, however, make women underperform compared to men (Gneezy, Niederle, and Rustichini 2003). This relative underperformance puts women at a disadvantage that might undermine the positive effects of increased participation. This disadvantage may extend to both being selected for a job and being promoted within jobs.

Gender differences in performance under competition are a well-established finding in the behavioral economics literature (e.g., Gneezy et al. 2003; Niederle 2016). Yet, this finding has barely been incorporated into the established sociological literature on gender inequality (e.g., Correll and Ridgeway 2006; Ridgeway 2014). Moreover, the dominance of economists in the gender and competition literature leaves an abundance of accumulated knowledge on the causes and consequences of gender inequalities unused. Sociology and social psychology have made major theoretical and empirical contributions in comprehending societal gender inequalities.

Because of the omnipresence of competition in modern societies, a better understanding of gender inequalities requires an interchange between these two strands of social science literature. Our exploration of the underlying causes of gender differences in competitive settings is guided by important insights from the sociological and social-psychological approach. In particular, we will argue that expectation states theory (Berger, Conner, and Fisek 1974), status characteristics theory (Correll and Ridgeway 2006), and the stereotype content model (Fiske et al. 2002) all have direct implications for understanding gender differences in competitive environment. In short, we intend to show that appreciating how competition affects men and women is important, but especially that understanding how situational contexts affect men and women under competition is essential for getting to the bottom of gender inequality.

There are various aspects of competition that may differentially affect men and women. To illustrate, think of job openings in academia. When a university posts an opening for a professorship, typically many candidates apply. This generates a competition for the position that has several dimensions. Two important dimensions are the following (Schram, Brandts, and Gërxhani 2019). First, there is a *rivalry for resources*. In this example, the (scarce) resource is a single position that is to be allocated amongst the candidates and that only one of them can obtain. Second, competition leads to a ranking of the competitors. This ranking remains private if only the candidates learn how they compare to (some of) the other applicants. In this case, we speak of a *private ranking* dimension of competition. The ranking becomes public if also other people know. Public ranking reflects a *social-*

status ranking dimension of competition (Ball et al. 2001).[1] The role of other people in social-status ranking is important. "Status refers to an individual's standing in the hierarchy of a group based on the prestige, honor, and deference accorded her by *other members*" (Lovaglia, Lucas, and Thye 1998, p. 202; italics added). The rivalry and ranking dimensions can be distinguished in most instances of competition. Either dimension or both can affect how people respond to competitive environments. If they are both at work, then their interaction may be important. In particular, they may either reinforce each other as complements or replace each other as substitutes.

Recognizing the importance of competition in modern societies and the increasing participation in competitive arenas by both men and women, we believe that a proper understanding of gender (in)equality requires more knowledge about how men and women are affected by the two dimensions of competition in isolation and by their interaction. We aim to contribute to this knowledge by theoretically and experimentally addressing two research questions. First, *are there gender differences in how people respond to rivalry for resources, to status ranking, and, in particular, to their interaction*? Second, *what mechanisms underlie such gender differences if they exist*?[2]

We rely on inputs from sociology and social psychology to motivate two principal mechanisms that can potentially explain the observed gender differences in performance under competition. These two mechanisms are the activation of (1)

[1] Henceforth, we will use the labels 'private ranking' and 'social-status ranking' when referring separately to the private and public aspects of ranking in competition, respectively. When referring more generally to the ranking dimension of competition, we will use the label 'status ranking'.
[2] When referring to a 'mechanism' we apply the concept as discussed by (i.a.) Kazdin (2007). A mechanism describes "the processes or events that are responsible for [a] change; the reason why change occurred or how change came about" (Kazdin 2007, p.3).

beliefs about gender differences in performance; and (2) a *warmth stereotype* that prescribes that women should take into account how their actions affect others.[3] We will argue that both mechanisms imply that gender differences in performance under competition are endogenous to situational contexts. That is, both mechanisms are products of the gender system under competition.

Our two research questions are important from a policy perspective. Since competition has more than one dimension, in different environments either dimension of competition may be more important or more salient than the other. Understanding the impact of the different dimensions of competition and their interaction will help in directing public policies aimed at reducing gender gaps to the dimension where their effect is deemed to be largest. Moreover, by identifying potential explanatory mechanisms, we hope to contribute to a better understanding of the *processes* that lead to gender inequalities under competition. This, too, intends to aid in the development of policies that may mitigate such inequalities. We make the link to policy in the concluding discussion.

GENDER AND COMPETITION IN ECONOMICS

In economics there is now an established strand of experimental research on gender differences in relation to competitive circumstances. Two related but distinct issues have been studied. First, starting with Gneezy et al. (2003) there is a stream that studies gender differences in performance under tournament incentives, that is, in an environment where only top performers obtain a scarce resource (as opposed to

[3] The warmth stereotype is also based on beliefs. As discussed below, however, it reflects a 'prescriptive' stereotype about how women ought to behave as opposed to a 'descriptive' stereotype of how they actually are (Prentice and Carranza 2002).

a situation where rewards are proportional to performance). It has been repeatedly observed that under such rivalry for resources women underperform relative to men. Second, following Niederle and Vesterlund (2007) there is a stream of work that studies gender differences in the disposition to work under competition, measured by the willingness to enter a tournament. These studies show that, compared to men, women tend to avoid competitive environments when they can. In both streams, the observed gender differences under competition appear to be quite robust. In her recent survey of the experimental literature in economics on gender, Niederle (2016) writes that the evidence for gender differences in competitiveness is more solid than that for gender differences in altruism or risk aversion (two other behavioral features that have been studied extensively).[4]

The literature on gender and competition in economics has been very influential in establishing the importance of competition for creating gender differences. It has, however, focused very strongly on one dimension of competition – rivalry for resources – and much less so on the effects of the ranking of individuals that competition involves. The effects of ranking only appear once in Niederle (2016). This is when she refers to Kuhnen and Tymula (2012) who document that giving relative performance information in the absence of rivalry for resources activates competitive attitudes and that women are more strongly affected. Various recent studies (Schram et al. 2019; Brandts, Gërxhani, and Schram 2020; Gërxhani 2020) show that social-status ranking – a performance ranking that is socially recognized – creates a stark difference in men's and women's performances in the absence of any rivalry for resources. They study performance in a real-effort task

[4] For earlier surveys, see Niederle and Vesterlund (2011) and Azmat and Petrongolo (2014).

involving numbers search and summation. When there is no social-status ranking, they find no gender differences in the number of attempted summations, nor in performance. By contrast, when there is social-status ranking men significantly increase the number of attempted summations as well as the number of correct summations. Remarkably, when women are subjected to social-status ranking, they significantly decrease the number of attempted summations. The net result is striking. With social-status ranking, men attempt more summations and correctly solve many (and highly significantly) more than women.

Previous research in economics tends to focus less on possible mechanisms behind observed behavior under competition. This may be due to economics having been traditionally more focused on (equilibrium) outcomes than on the processes leading to outcomes. In the next section we draw on some important insights from sociology and social psychology and discuss how these insights allow us to delineate the mechanisms that help us better understand gender differences in performance under competition.

GENDER INEQUALITY IN SOCIOLOGY AND SOCIAL PSYCHOLOGY

Social Relational Contexts and Commonly Shared Gender Beliefs

Sociological theory has generated several concepts that are useful in building a broader framework for studying competitive environments. We would argue that competition creates a 'social relational context', a setting extensively studied by Ridgeway and Correll (2004). This is defined as follows: "Social relational contexts comprise any situation in which individuals define themselves in relation to others in order to act." (Ridgeway and Correll 2004, p. 511). In such a context, expectation states theory and its best-known branch status characteristics theory apply (Berger,

Wagner, and Zelditch 1985; Correll and Ridgeway 2006). Traditionally, expectation states theory has focused on social relational contexts in which individuals are oriented toward accomplishing a collective goal (Wagner and Berger 1997), that is, in 'collectively oriented task groups' (Berger et al. 1974). These include most work and educational contexts but also many informal and personal goal-oriented contexts. The theory argues that, when gender is effectively salient in such settings, beliefs about men's greater competence and status can implicitly shape the expectations that participants form for their own competence relative to that of other group members in the setting.

By now, numerous studies have shown that status hierarchies and their implications for individuals' performances are also present in 'individual evaluative tasks' without collective goal (Foschi, Lai and Sigerson 1994; Erickson 1998, Lovaglia et al. 1998; Correll 2001). Whenever individuals feel that they will be socially evaluated, they may experience some pressure to assess their competence relative to others "who they imagine are also being or have been evaluated" (Correll and Ridgeway 2006, p. 47).[5] This social comparison leads to a (possibly implicit) ranking of expected performance. When objective information on one's (relative) competence is lacking –and at times even when it is known (Foschi et al. 1994)– prominent characteristics such as gender, race or age may be used as a 'status characteristic' where one category (e.g., men) is believed to perform better than the other (women). Individuals may thus resort to a status characteristic like gender even when the task at hand is individualistic, as long as there are commonly shared

[5] This is related to the views of symbolic interactionists, who argue that for individuals acting alone the social environment is still highly relevant if they expect an evaluation of their performance (e.g., Stryker and Vryan 2006).

gender beliefs that one gender is generally more competent and has a higher status than the other.

As argued by status characteristics theory, the more 'hegemonic' (Ridgeway and Correll 2004) the gender beliefs are, that is, the more widely they are shared across a society, the stronger will be the role of status hierarchies based on gender. Irrespective of actual abilities, hegemonic gender beliefs can change how people see their own ability and subsequently how they perform (Biernat and Kobrynowicz 1997; Spencer, Steele, and Quinn 1999; Foschi 2000; Correll 2004; Ridgeway and Correll 2004). When it is a priori ambiguous what constitutes a 'good' performance and it is also uncertain how others perform, hegemonic gender beliefs provide a benchmark for expectations about own performance. As a consequence, once status hierarchies have been established, they can have a strong influence on performance in individual tasks (as confirmed experimentally by Lovaglia et al. 1998).

In short, when a status characteristic like gender is salient (i.e., easily perceivable) and men are generally believed to outperform women, then even in individual evaluative tasks, men are predicted to indeed outperform women. Moreover, both female performers themselves and outside performers will consider women less 'able', even if they perform equally to men (Correll and Ridgeway 2006, p. 47). Note also that such beliefs –while suppressing female performance– may have the opposite effect on men, making them believe to be better and boosting their performance (Correll 2001).

These predictions are highly relevant for gender differences in performance under competition. Status characteristics theory can be directly applied to the two dimensions of competition. For status-ranking, this is perhaps clearest. Social-status ranking implies an explicit social evaluation (Lovaglia et al. 1998) while such an

evaluation is implicitly made by an individual in private ranking (Gërxhani 2020). As for the rivalry-for-resources dimension, the social evaluation involves a dichotomous ranking dividing those who obtain the resource from those who do not. Hence, when competition involves either a gendered salient task or a mixed-gender setting where a comparison with the other gender category becomes salient, gender is expected to come into play as a status characteristic. Indeed, a salient gender task and a mixed-gender setting have traditionally been considered as two important scope conditions for status characteristics theory to apply (Ridgeway and Correll 2004). Under these conditions, hegemonic gender beliefs evoke 'stereotype threats' (Steele 1997) and 'social evaluative threats' (Dickerson and Kemeny 2004) that may negatively affect women's performance. We propose that such beliefs may also be activated by another characteristic of the social relational context. This is the context of competition *per se*, which may serve as an "environmental trigger" (DiMaggio 1997, p. 279) that activates gender beliefs and stereotypes.[6]

We thus expect social evaluative threat to arise in competitive settings. This is because one is explicitly compared to others (the status-ranking dimension of competition) and because one ends up either as a 'winner' or as a 'loser' of the competition (rivalry for resources). In comparison to performing in a non-competitive environment, the heightened state of social evaluative threat that comes with competition per se could make men and women more likely to act on the beliefs that they hold. In other words, the hegemonic gender beliefs that men are better than women will be more strongly activated in a competitive environment, which in turn will affect individual performances. In this way, gender may come into

[6] See Castilla and Benard (2010) for a similar line of argument about meritocracy.

play as a status characteristic in the social relational context of competition because of the task being competed on, because of a mixed-gender environment, or because of the competition per se.

In our experimental studies, we create distinct environments (or "arenas") in which individual tasks are conducted. These environments differ in the competitiveness of the relational contexts involved. This distinction allows us to vary the extent to which gender beliefs are activated by competition per se, and thus, how competition may impact men's and women's performances (Deaux and LaFrance 1998; Ridgeway and Smith-Lovin 1999). When the context involves no explicit comparison with others' task performance, cultural gender beliefs are expected not to be explicitly activated. Nevertheless, the mere fact that one knows that others have done, are doing, or will do the same task may *implicitly* activate such beliefs (Ridgeway and Correll 2004). This has the following consequences for our experimental setting. With competition, the environment we create involves *explicitly* comparing individual task performances. Without competition, they may be implicitly activated. We expect cultural gender beliefs to be more strongly activated in the former case.

Beliefs about gender differences play a key role in our argument that competition causes gender differences in performance. Indeed, as discussed above, gendered beliefs are one of the mechanisms in our theoretical understanding of this causal relationship. The next section provides a more detailed discussion of our mechanisms and their role in explaining the relationship between competition and gender differences in performance.

Social Relational Contexts and the Warmth Stereotype

Aside from gendered beliefs related to competence, a social relational context (like competition) may also activate other stereotypical behavior. In particular, while men are perceived to be more competent, the 'stereotype-content model' in social psychology argues that women are expected to be 'warmer' than men (e.g., Ebert, Steffens, and Kroth 2014).[7] In this context, warmth refers to being, e.g., empathetic, good natured, sincere, and caring (Ebert et al. 2014; Connor and Fiske 2018); a stereotype that originates from traditional gender roles. Importantly, such a stereotype is not descriptive of how women behave *per se*,[8] or even how women are believed to be. Instead, this stereotype *prescribes* expected gender behavior both in terms of what women should and should not do in certain situations. As described by Heilman and Okimoto (2007, p. 81):

> "The dictates of prescriptive sex stereotypes are highly specific and widely shared. They specify that women should behave communally, exhibiting nurturing and socially sensitive attributes that demonstrate concern for others, such as being kind, sympathetic, and understanding. They also specify what women should not do—engage in behaviors typically prescribed for men that are thought to be incompatible with the behaviors prescribed for women.

[7] Fiske et al. (2002) allow for various mixed stereotypes beyond the warmth-competence dichotomy. As will become clear from the arguments that follow, the warmth-competence distinction suffices to explain the effects of gender differences in competition.

[8] The empirical evidence is mixed on whether women are warmer towards others than men. On the one hand, Stuijfzand et al. (2016) find in two observational studies that female adolescents are more empathic than male adolescents and Willer, Wimer, and Owens (2015) report that men give less to poverty relief than women do (and attribute this to gender differences in levels of empathy). On the other hand, in an extensive survey of the experimental literature, Niederle (2016) finds no evidence of a gender difference in altruism. This mixed evidence reinforces the idea that such gender differences may be situationally dependent.

> Thus, agentic behavior, behavior that demonstrates dominance, competitiveness, and achievement orientation, is generally considered out of bounds for women."

The latter prescription implies that women should avoid success in competition and should instead exhibit communal behavior.

It follows that gender differences in warmth are not to be considered stable behavioral tendencies that are exogenous to situational contexts. Women may, however, show more warmth than men because it is expected from them in specific interactions. In fact, women may expect sanctions, retribution, conflict, or diminished likeability for 'cold' behavior, such as performing relatively better than others (e.g., Heilman et al. 2004). Combining an audit study with a survey experiment, Quadlin (2018) finds that competence and commitment are highly valued by employers when considering male applicants, while likeability is perceived more important for female applicants. As a consequence, high-achieving women are viewed with skepticism, while sociable moderate-achieving women are highly rated. Bursztyn, Fujiwara, and Pallais (2017) report similar results for the 'marriage market'. They find that three-quarters of single female students at an elite US MBA program report having avoided activities they thought would help their careers to prevent looking ambitious, assertive, or pushy. They are more likely to have avoided these activities than non-single women or men. For more related evidence, see also King et al. (2017) and Gino, Wilmuth, and Brooks (2015).

The question, then, is how someone who intends to exhibit warmth behaves in a competitive environment. Because one's standing in a competition is relative, competitive success imposes costs on others (Frank, 2004; Willer et al. 2013). 'Warm' behavior in competition then involves taking into account the costs that

success invokes in one's competitors. In particular, a warm response would involve avoiding these costs by reducing one's own chances of competitive success.

In summary, an environment where gender is salient may yield hegemonic stereotype beliefs that women *should* behave in a way that exhibits warmth. Such beliefs are reinforced by sanctions when women do not behave accordingly. In turn, displaying warmth may be an important attribute that affects behavior in a competitive environment. In this line of reasoning, relative to men, women may underperform in competition in order to diminish the costs to others.

EXPLANATORY MECHANISMS

A variety of supply- and demand-side explanations have been advanced in the literature to understand gender differences in the access to high-level positions in society.[9] Demand-side factors are barriers that hinder women's access to high positions, often related to different kinds of discrimination (e.g., Heilman and Parks-Stamm 2007; Neumark 2018). Supply-side factors are differences in perceptions held, decisions made, or behaviors enacted by men and women themselves, 'whether free or constrained', that contribute to gender differences (Ceci and Williams 2010; Gino et al. 2015).

Per design, our focus is on supply-side explanations, and in particular those that affect performance. Based on the literature discussed above, we consider two mechanisms that potentially influence behavioral outcomes under competition in our experimental studies. Aside from investigating directly whether the mechanisms are at play, we also dig deeper by analyzing in more detail how they

[9] See Gino et al. (2015) for a summary of the most-heard demand- and supply-side explanations.

operate. We do not claim this to be an exhaustive list of supply-side factors. For example, fertility choices, work-home balance, career preferences and ability differences have all independently or in combination been shown to be important for understanding gender differences in performance (e.g., Ceci and Williams 2010). Our choice of mechanisms is guided on the one hand by the existing theoretical and empirical knowledge of the relationship between competition and gender differences and on the other by a desire for parsimony in the method we apply.

To start, Figure 1 summarizes the role of the two principal mechanisms that we consider, based on the sociology and social psychology literature discussed above. Both mechanisms describe direct causal links from competition to gender differences in performance. The importance of these mechanisms for understanding gender inequality is well established in sociology. Our contribution lies in showing how they are also activated under different competitive settings and how they explain the gender differences in competitive settings that have been primarily studied in economics.[10]

<Figure 1 about here>

The first mechanism, which we denote by M1, is based on the performance-beliefs-activation arguments provided above. If competition per se serves as an environmental trigger, then the degree to which beliefs about expected gender differences (related to a particular task) will be activated will vary with the competitiveness of the environment. To investigate this possibility, we will elicit beliefs under no competition and under diverse competitive environments. Note

[10] Such mechanisms are typically not studied in economics. One exception is a recent working paper by Buser, Cappelen, and Tungodden (2021). They study the role of fairness concerns in the willingness to compete. They find that fairness considerations cannot explain why women (relative to men) shy away from competitive environments. They do not, however, study the effects of these considerations on performance.

that at this stage, we do not know how either or both dimensions of competition affect gender beliefs. For this reason, we will elicit such beliefs separately for both rivalry for resources and social-status ranking. If either of these competitive environments triggers a social evaluative threat where gender is salient, we should observe stronger beliefs that women perform more poorly than men under one or both dimensions of competition than without competition. In turn, this will yield larger gender differences in performance, given the feedback from beliefs to behavior.

The second mechanism (M2) builds on the observation that competition generates an environment where a good performance has a negative impact on others, that is, there is a negative externality of good performance. As argued above, competition may then activate a prescriptive gender stereotype of female warmth, where women are expected to show concerns about how their performance affects others. We will test this mechanism by creating an environment where the competition is with others who have already finished the competition in the past. In this setting, one's own competitive success has no consequences for others and concerns for others should play no role.

To further understand how these mechanisms operate, we explore two additional questions –depicted in Figure 2– which builds on Figure 1.

<Figure 2 about here>

The first question, shown in the top panel, examines whether the process through which competition activates gendered performance beliefs and a prescriptive warmth stereotype –which consequently lead to gender differences in performance– is moderated by the *gender composition of the group of competitors*. As mentioned above, cultural gender beliefs are most expected to play a role in mixed-gender settings. This is because individuals in mixed groups can compare

themselves to the opposite gender, thus activating widely shared beliefs about gender differences in performance (Ridgeway and Correll 2004). Similarly, prescriptive stereotypes about female warmth may not be activated in single-gender environments. As proposed above, however, such beliefs and stereotypes might also be activated and therefore affect performance by competition *per se* if the anticipation of rivalry or status ranking that this involves serves as an environmental trigger. In other words, even if competition is with others of the same gender, the mere fact that one is competing might generate M1 and M2. Though it is *a priori* unknown how the social evaluative threat caused by competition *per se* relates to that caused by a mixed gender environment, we intuitively expect that the mechanisms will be stronger when competition and mixed gender are combined than when competition is with others of the same gender.

If gender composition moderates the activation of gender beliefs and stereotypes under competition, women will underperform (relative to men) more in a mixed-gender than in a same-gender competition. Indeed, studying rivalry for resources, Gneezy et al. (2003) find evidence of reduced gender differences under same-gender competition than when competition involves both men and women. Also, Niederle and Vesterlund (2010) argue that girls' performance on math tests and their willingness to compete in high-stakes testing environments are influenced by the gender of the other competitors and test takers. This is confirmed in a high-stakes field setting by Van Dolder, Van Den Assem, and Buser (2020). We will investigate whether these findings replicate in our data and whether they also hold for the status-ranking dimension of competition. Note that in this theoretical reasoning, we assume that the gender composition under competition affects performance by moderating the activation of mechanisms M1, gendered

performance beliefs, and M2, prescribed warmth stereotype. Empirically, we will test whether the gender composition of the group of competitors affects gender differences in performance; we represent the intermediary position of beliefs and stereotypes by the dashed-line box in the top panel of Figure 2.

The second additional question that we explore to further understand how the two principal mechanisms operate is depicted in the lower panel of Figure 2. We ask whether the gender effects of competition on performance (via mechanisms M1 and M2) are caused by differential effects on effort. The idea here is that either of the two mechanisms may demotivate women relative to men (Correll 2001, p. 1699; Correll and Benard 2006), which may make them reduce their efforts and subsequently lead to gender differences in performance. Such demotivation was observed in a recent study, that tries to understand the gender gap in the highly competitive STEM fields (Penner and Willer 2019). The study argues that not only do many women 'under-persist' by failing to pursue careers in science and mathematics despite sufficient qualifications, men also often make an extra effort and 'over-persist', by choosing STEM even when it might lead to less success than non-STEM options would. A possible explanation for this discrepancy originates in the demand side. In STEM fields high-achieving women are considered less committed (Rivera 2017) or less 'likeable' (Quadlin 2018). Competition may play an important role in this line of argument. It is precisely in highly-competitive environments that men are believed to outperform women and that men who excel are considered differently than women who excel. In short, competition may make women feel demotivated or withdrawn due to expected poorer performance or negative consequences of performing better than men. Such a response to these beliefs and

stereotypes will make women put in less effort than men. In turn, these effort responses could yield a gender gap in performance and thereby contribute to gender inequality.

FIVE STUDIES

Our research questions, the two principal mechanisms, and their elaboration lead us to conduct five distinct studies. Study 1 delineates the impact on performance of the rivalry and status-ranking dimensions of competition and their interaction. The focus in Study 2 is on the first mechanism (M1), where we measure 'beliefs' about gender differences in performance. Study 3 explores the second mechanism (M2), 'prescribed warmth stereotype'. Studies 4 and 5 consider the further questions depicted in Figure 2, with Study 4 addressing whether observed performance differences depend on the gender composition of the environment and Study 5 investigating whether the process through which the two mechanisms affect performance is channeled through gender differences in effort. We discuss each study in turn.

STUDY 1 – DISENTANGLING THE DIMENSIONS OF COMPETITION

Experimental Design

For study 1 we conducted sessions with six treatments involving performance on a cognitive task under different combinations of rivalry and ranking. In each of the six treatments there are two parts which involve three types of subjects: *A*s, *B*s and *C*s. In most sessions there are six *A*s, six *B*s and one *C*.[11] Instructions for each type are

[11] As will be explained below, there were also sessions that did not require A or C players; in these sessions, there were two groups of six *B*s.

presented in Appendix A. In Part 1, *A*s and *B*s work individually on a cognitive task and *C*s are passive. Treatments vary in the monetary incentives *A*s and *B*s face and in the information they receive. This is explained below.

The cognitive task in part 1 is the same in all sessions and is taken from Weber and Schram (2017).[12] Participants are presented with a sequence of pairs of 10x10 matrices filled with two-digit numbers. These matrices appear at the lower half of their computer monitor (Figure 3). For each pair of matrices each participant is asked to individually search to find the highest number in the left matrix and the highest number in the right matrix and to calculate the sum of these two numbers. This sum must be entered in the window at the center-top of the monitor. After a number has been entered, two new matrices appear, regardless of whether the sum was correct or not. The task continues for 15 minutes. At any time during work on the task each participant can see on the screen the own cumulated number of correct summations as well as the remaining time. Participants are not informed about the performance of any other participant.[13]

<Figure 3 about here>

The instructions emphasize the importance of doing well in this task by mentioning that it has been shown to correlate positively with success in professional life.[14]

[12] Alternatively, we could have used the summation task applied in Niederle and Vesterlund (2007). Shurchkov (2012, fn 21), however, reports evidence of a stereotype threat in this task *per se*, where women feel a priori that men have an advantage. To avoid this, we decided to use a task that we have applied before (Weber and Schram 2017; Schram et al. 2019). In these previous studies there was no evidence of gender differences and our data for B-players in this Study 1 confirm this. This is why we believe there to be no stereotype threat for the task per se. This result allows us to focus on the effects of competition on gender inequality, without needing to deal with noise from the task itself.

[13] Our participants perform a so-called real-effort task (as opposed to stated effort, which is often used in experimental work). Real effort is a necessary component of our design, *i.a.,* because we consider exerted effort as one of our mechanisms.

[14] This emphasis was made to stress the importance of status ranking based on the performance in the particular task we used. Additional tests reported in Schram et al. (2019) show that this priming does not induce stereotype threat. In fact, excluding it does not significantly affect results. Nevertheless, we include it here to maintain consistency across related projects.

Participants were told that we would provide evidence of this claim upon request after the experiment. For this purpose, we had available copies of Koedel and Tyhurst (2012), a résumé study linking math skills to labor market outcomes that provides such evidence.

The way in which participants are rewarded for task performance is one of our treatment variables. In a similar vein to Gneezy et al. (2003), we use two different payment schemes. The first is an individual piece-rate payment, with each correct answer yielding one euro. The second is a tournament payment scheme, where only the two participants with the highest score in a group of six receive payment for each correct answer, while the other four receive nothing. To keep average payment constant across treatments the payment per correct answer with tournament payment is three times that of piece-rate payment: three euros. The idea underlying this treatment variation is that the tournament payoff creates a rivalry for resources, while the piece rate does not. For this reason, we use the acronyms *nRfR* (no rivalry for resources) and *RfR* (rivalry for resources) for the piece-rate and tournament incentive treatments, respectively.

Our second treatment variation is used to study the status-ranking dimension of competition, building on the design of Schram et al. (2019). This is applied in part 2, where we vary whether or not participants receive ranking feedback. One group of *B* players receives no such feedback; we call this the no-Ranking treatment (*nR*). This means that these *B* players are passive in part 2; the *C* player also remains passive in part 2. Other players do receive feedback, which may be one of two types. Recall that status ranking has two distinct characteristics. It informs an individual of her own ranking vis-à-vis others and it informs others of her ranking. By varying the feedback participants receive, we isolate the former.

This allows us to differentiate between the effects these characteristics might have. In the first type of feedback participants are only informed about their own ranking. In particular, in some sessions, a different group of B-players than those in *nR* is privately given this information. We call this the private-ranking treatment, *PR*. The second type of feedback is provided in the social-status ranking treatment (*SR*), which involves the following. After finishing the instructions, each *A* player is individually taken to the *C* player, who does not take part in the real-effort task and whose task consists in listening. The *A* player reads aloud a text stating that she will return after the task to report her performance score. This is done to create the anticipation of having to later report to the *C* player. After finishing the task, the *A* player is informed of her score and rank amongst the group of six *A* players. She is then individually taken to the *C* player a second time, where she reads a text stating her score and rank. The *C* player, a peer, is the same for all *A*s in a session, so that this person will end up knowing the rank of each of the participants in the real-effort task. In all cases, the ranking condition is common information. Importantly, both private and social-status ranking information consist in knowing one's own position in the ranking, but not the complete ranking of all relevant participants. The only person who has this complete knowledge is the *C* player.

We crossed tournament pay and piece-rate pay with the three ranking treatments in a full-factorial design, yielding the total of six treatments. Table 1 summarizes these treatments.

<Table 1 about here>

Procedures

Study 1 was run at the BLESS laboratory of the University of Bologna, Italy in 34 sessions in the period November 16 - December 7, 2017.[15] 442 participants were recruited from the BLESS participant pool for sessions that lasted 60-75 minutes. Average earnings were approximately 18 euro.

Testable Hypotheses

Our first goal is to study the relative importance of rivalry for resources and social-status ranking and their interaction. For this purpose, we combine the designs of Gneezy et al. (2003) and Schram et al. (2019). As a consequence, we expect to replicate the Gneezy et al. (2003) results on the effects of rivalry for resources when there is no status ranking (*RfR*/*nR*) and, *vice versa*, we expect to replicate the Schram et al. (2019) results on the effects of status ranking when there is no rivalry for resources. This yields, respectively, Hypotheses 1.1 and 1.2. Preceding these, our first hypothesis (1.0) is that there are no gender differences in performance if neither of the two dimensions of competition is active (*nRfR/nR*). This assumes that there is no stereotype threat related to the task per se, that is, that gender beliefs and stereotypes are not activated when individuals do the task in isolation. This assumption is based on the benchmark results summarized in Gërxhani (2020).

[15] We thank Marco Casari for making the laboratory available and Lorenzo Golinelli, Mario Spiezio, and Giorgio Monti for their assistance in organizing the experiments.

Hypothesis 1

1.0: Without rivalry for resources and without either kind of status ranking, men and women perform equally. In other words, no gender differences in performance are expected in the treatment combination *nRfR/nR*.

1.1: Without either kind of status ranking, men perform better than women under rivalry for resources. In other words, men are expected to perform better than women in *RfR/nR*.

1.2: Without rivalry for resources, men perform better than women under either kind of status ranking. In other words, men are expected to perform better than women in *nRfR/PR* and *nRfR/SR*.

An important part of our research questions concerns the interaction between rivalry for resources and the social-status ranking dimensions of competition. A priori, we can conceive of no solid theoretical basis for predicting the interaction, nor do we have previous results to rely on. We can think of two opposing ways in which this interaction may take place. We illustrate this for mechanism M1 (gendered beliefs). If the effects of the two dimensions are completely separated, then we should consider them as complements. In that case, gender beliefs are already activated if there is only status ranking[16]. They are activated more strongly, however, if rivalry for resources is added. With complements the same holds if status ranking is added to a pre-existing rivalry for resources. As a consequence, the combination of rivalry for resources and status ranking will yield a larger gender difference than either dimension of competition alone. On the other hand, both

[16] For ease of presentation, we do not distinguish here between private- and social-status ranking. When formulating our empirical question below, we will make this distinction.

dimensions of competition might have the same effect on men and women but without reinforcing each other. In other words, the two dimensions might act as substitutes. In the case of gendered beliefs, this would mean that these are already activated when one of the dimensions is in place, while adding the other dimension does little to strengthen these beliefs. There is then somehow a 'ceiling' for gender performance differences.[17] Because we have no hypothesis about whether the competitive dimensions are complements or substitutes, we pose the following empirical questions:

Empirical Questions 1.

1.1 Are private ranking and rivalry for resources complements or substitutes? In other words, are gender differences in *RfR/nR* similar to or smaller than in *RfR/PR* and are gender differences in *nRfR/PR* similar to or smaller than in *RfR/PR*?

1.2 Are social-status ranking and rivalry for resources complements or substitutes? In other words, are gender differences in *RfR/nR* similar to or smaller than in *RfR/SR* and are gender differences in *nRfR/SR* similar to or smaller than in *RfR/SR*?

[17] To illustrate, assume that there is no performance difference without any competition and that men score x units better than women under dimension 1 and y better under dimension 2. Assume without loss of generality that $y > x$. If the two dimensions are perfect complements, then adding dimension 2 to dimension 1 or vice versa increases the gender difference to $x + y$. If they are perfect substitutes, then adding dimension 2 to dimension 1 increases the gender difference from x to y while adding dimension 1 to dimension 2 keeps the gender difference at y.

Results

We collected data for Study 1 in 34 sessions with 13 participants each, yielding a total of 408 active participants (*A*s and *B*s) and 34 passive *C* players. We used power analysis to determine the desired numbers of observations. Based on the social-status ranking treatment in Schram et al. (2019), a power of 80% and significance level 5% would require a sample of 12 male and 12 female participants if means are tested with a t-test. We have more treatments than just *SR*, of course and therefore need more observations. On the other hand, unless indicated otherwise, we use permutation t-tests using Monte-Carlo resampling with 10,000 repetitions (hence-forth, PtT) to compare means. This non-parametric Fisher-Pitman test has substantially higher power than standard t-tests (cf. Appendix C in Schram et al. 2019) and therefore requires fewer observations. In the end, we decided to stay on the safe side of these countervailing considerations and aimed at a minimum of 25 male and 25 female participants per treatment. Table 2 offers an overview of the numbers of male and female participants we ended up with in each treatment.[18] It also defines four outliers that were excluded from the analyses and shows the treatments they were in.

<Table 2 about here>

Figure 4, below, shows average performance for men and women in each of the six treatment combinations of Study 1. We use this to first consider hypothesis 1.0, which predicts no gender differences in *nRfR/nR*. The figure indeed shows only a small performance difference (the difference is 1.2 correct summations in 15 minutes). This difference is statistically insignificant (PtT; $p = 0.260$, $N = 60$) when

[18] To avoid signaling our interest in gender differences, we did not recruit separately per gender. For this reason, the realized numbers show some random fluctuation, but the minimum of 25 observations per gender was realized in all cases.

testing a null of no gender difference against an alternative hypothesis that a gender difference exists. Note, however, that our hypothesis 1.0 does the opposite; it has an alternative hypothesis of no gender difference. This invalidates testing against a null of no difference (e.g., Raftery 1995). Finding support for the hypothesis requires testing our hypothesis of no difference against a hypothesis where there is a difference. To do so, we apply Bayesian testing (Berger 2013). In particular, we compare two models that predict outcomes in *nRfR/nR*. One model is based on our hypothesis that there is no gender difference in performance. The other model predicts that behavior in *nRfR/nR* is similar to the behavior observed in treatment *nRfR/SR* (where a gender difference is observed, as discussed below). Bayesian analysis then allows us to determine the likelihoods that the behavior that we observe in *nRfR/nR* can be attributed to either of the two models. The results show that 'our' no-difference hypothesis is three time more likely to have generated the observed behavior than the *nRfR/SR*-based model. This provides evidence in favor of hypothesis 1.0. More details are presented in Appendix B.

To investigate hypotheses 1.1 and 1.2, Table 3 shows the mean performance difference between men and women in the various treatments and the *p*-values for PtT tests on the null that there is no gender difference.[19] This shows that the gender differences in the other five treatments are much larger than in *nRfR/nR*, and that men perform significantly better than women in all these cases.[20]

<Figure 4 about here>

<Table 3 about here>

[19] As just discussed, for the *nR* case the more appropriate statistical procedure is Bayesian. Nevertheless, for comparability, Table 3 also shows the result of the PtT for *nR* (p=0.256).

[20] Below, we will investigate whether the observed gender differences under competition can be attributed to women underperforming, men overperforming, or both. Here, we follow much of the literature in focusing on the net effect: the gender difference in performance within each treatment.

For now, we disregard the last two columns of Table 3 and focus on the cases where the two dimensions of competition are not simultaneously active. Hypothesis 1.1 predicts that men perform better than women when there is only rivalry for resources (*RfR/nR*). We find strong evidence in favor of the hypothesis. This replicates the results first observed by Gneezy et al. (2003) that rivalry for resources makes women underperform in comparison to men. Hypothesis 1.2 predicts that men perform better than women when there is only (private or social-status) ranking (*nRfR/PR* and *nRfR/SR*, respectively). Again, both predictions find strong support in our data. The effect for *nRfR/SR* replicates the effects of social-status ranking reported in Schram et al. (2019). Here, we observe a significant gender difference even when participants are only privately informed about their ranks (*nRfR/PR)*, though this difference is smaller (2.71) than when a peer is informed about this ranking (3.45). We show in Appendix B that this difference-in-difference (2.71 vs. 3.45) is statistically insignificant. All in all, we find support for each of the hypotheses 1.0–1.2.

Result 1a: Our data support the hypothesis that there is no gender difference without competition of any kind (Hypothesis 1.0). When there is either rivalry for resources or (private or social) status ranking, men perform significantly better than women (Hypotheses 1.1 and 1.2).

Now consider the joint effects of rivalry for resources and status ranking, that is, treatments *RfR/PR* and *RfR/SR,* on performance. This joint occurrence of two dimensions of competition has, to the best of our knowledge, not been previously studied explicitly. Notice in Table 3 that the gender difference is highly significant in both cases when testing against a null of no gender difference. Empirical Questions 1.1 and 1.2 ask whether or not the gender difference when both dimensions of

competition are active is different than when participants only face one dimension, that is, whether the dimensions are complements or substitutes. To address this, we again adopt a Bayesian approach.[21]

The substitutes model predicts that the gender difference observed when only one dimension of competition is active does not change if a second dimension is added. Note, however, that in the data the gender difference in (only) one dimension might be different than in (only) the other. For example, Table 3 shows that the difference is 3.63 when there is only rivalry (*RfR/nR*) and 2.71 when there is only private ranking (*nRfR/PR*). When both dimensions are active (*RfR/PR*) the difference is 4.89. The substitutes model predicts no change if one adds the second dimension; it compares the two-dimension case *RfR/PR* to a case with only one competitive dimension. Because the latter is either 3.63 or 2.71, however, it matters which one-dimension case one compares too. We solve this by making both comparisons. In other words, we compare *RfR/PR* to *RfR/nR* (4.89 to 3.63) and to *nRfR/PR* (4.89 to 2.71), separately. We refer to the treatment that *RfR/PR* is compared to as the "baseline comparison". There are then two baseline comparisons for *RfR/PR*, to wit, *RfR/nR* and *nRfR/PR*.

In the complements model, the gender difference increases when adding a second dimension. Once again, we need to specify the baseline comparison that is made. Here, we also need to predict by how much the gender difference will increase. To do so, we take as a benchmark a comparison to the case without

[21] In particular, we compare the probability (1) that the substitutes model explains our data, to the probability (2) that the complements model explains the data. Dividing the probability (1) by (2) gives the so-called odds ratio. An odds ratio of 10/1 implies that the substitutes model is 10 times as likely to have generated the data than the complements model, while an odds ratio of 1/10 implies the reverse.

competition, *nRfR/nR*. By comparing the case with a single dimension of competition to this non-competitive treatment, we obtain a benchmark prediction of how this single dimension affects gender differences. The model of complements predicts that the gender difference will increase by the same amount when adding that dimension to the another. For example, the gender difference without competition (*nRfR/nR*) is 1.2, and with only rivalry (*RfR/nR*) it is 3.63. The benchmark prediction for the effect of rivalry is then 3.63 – 1.2 = 2.43. The complements model thus predicts that the gender difference is also 2.43 larger in *RfR/PR* than in *nRfR/*PR.

The preceding description summarizes the general procedure that we used to study empirical questions 1.1 and 1.2. The Bayesian analysis requires a statistical analysis that assigns probabilities to the models' predictions. For example, by properly taking into account that the benchmark prediction for the effects of rivalry is measured with noise and that predicted behavior in *RfR/PR* is stochastic, one can determine the likelihood that the complements model is correct. A more detailed description of the statistical implementation is presented in Appendix B. Table 4 summarizes the results.

<Table 4 about here>

The results for combining rivalry for resources with private ranking do not provide much support for either model relative to the other. For example, when adding rivalry to private ranking (column 2), a model that assumes that they are complements is more than twice as likely to be correct. For the reverse (adding private ranking to rivalry) both models are more or less equally likely. The evidence for combining rivalry for resources with social-status ranking, however, strongly favors the model of substitutes. In both cases (columns 4 and 5) this model is more

than 24 times more likely to be correct than a model where they are complements. This gives our next result.

Result 1b: Neither the complements model nor the substitutes model is favored when combining private ranking and rivalry for resources. Social-status ranking and rivalry for resources are much more likely to be substitutes than complements.

Conclusions

Results 1a and 1b establish that the task we use does not generate gender differences in performance when done in a non-competitive setting; that either dimension of competition does generate such gender differences; and that rivalry for resources and social-status ranking are substitutes. Relating these results to the mechanisms discussed above allows us to conclude that gender beliefs and stereotypes seem to be activated and lead to gender inequality in performance only when explicitly competing with others. Moreover, we add a novel contribution to this discussion, namely that beliefs and stereotypes activated by the two dimensions of competition are similar and do not reinforce each other. We now turn to a more detailed analysis of these mechanisms.

MECHANISMS

Study 2 – M1: Beliefs

The first mechanism that we consider (M1) is that competition activates *beliefs* about how men and women are expected to perform in the cognitive task.[22] The previous literature has discussed not only a role of beliefs, but also the activation of these beliefs by the social relational context. As we argue above, gender beliefs might reflect the task being competed on or the gender of the people one is competing with, but they might also be activated by an environmental trigger such as competition itself. Investigating whether beliefs can be a mechanism behind the performance differences reported in Study 1 then requires eliciting such beliefs in relation to the specific competitive conditions under which performance takes place.

In three new sessions, we therefore elicited beliefs about gender differences in performance in the absence of any competitive setting (nRfR/nR), under (only) social-status ranking (*nRfR/SR*), and under (only) rivalry for resources (*RfR/nR*). These sessions were run on December 10-11, 2019 at the BLESS laboratory in Bologna, Italy. We recruited 96 participants (48 male, 48 women)[23] for four sessions that lasted 50-60 minutes each. Average earnings were €17.82.

Each session consists of four parts. In the first, participants did the summation task with piece-rate remuneration of €0,50 per correct answer. This allowed them to get acquainted with the task. In the following three parts, we explained that previous participants (i) had done the same task for €1,00 per correct answer; or (ii) had done the same task for €1,00 per correct answer and then had to report their

[22] Though our focus here is on the supply side of the labor market, it is interesting to note that beliefs may also affect gender discrimination at the demand side of the labor market (Coffman, Exley, and Niederle 2019, Sarsons et al. 2021).
[23] It is a coincidence that the numbers of female and male participants are equal.

rank to a peer; or (iii) had done the same task and received €3,00 per correct answer (only) if they were in the top two in a group of six.[24] In each part, they were asked to predict whether the mean score of men was better than that of women or vice versa. We excluded the possibility of predicting exactly equal mean scores because this is an event with extremely low probability. One of these three parts was randomly chosen at the end of the experiment and every participant that had predicted correctly in that part received an additional payoff of €5.00.

As a measure of beliefs, we use the fraction of participants (per gender) that predict that women will perform better than men. To derive testable hypotheses for these sessions, we assume that gender beliefs are not systematically activated for the case where the task is done in isolation. In other words, we assume that there are no descriptive gender stereotypes for this task. As mentioned above, this assumption finds support in the benchmark results summarized in Gërxhani (2020). Any belief that women do better than men or *vice versa* is then driven by individual idiosyncrasies and roughly half the participants are expected to predict that women will do better. Under either dimension of competition, however, we assume gender beliefs to be activated and we expect that less than half of the participants will believe that women will do better than men. This gives:

Hypothesis 2

2.1: 50% of the women and 50% of the men believe that women perform better than men when no competitive dimension is active.

[24] To avoid order effects, each of the four sessions had a different order of (i), (ii), (iii) in parts 2-4. The order does not affect the reported beliefs, so we pool the data in our analysis. More details are available upon request.

2.2: Less than 50% of the women and less than 50% of the men believe that women will perform better than men when the task is done under social-status ranking.

2.3: Less than 50% of the women and less than 50% of the men believe that women will perform better than men when the task is done under rivalry for resources.

Table 5 summarizes the elicited beliefs by showing the fraction of participants that think that women perform better than men in the environment concerned. Without competition, it holds for both women and men that the proportion that believes that women are better does not differ significantly from 0.5 (as indicated by the binomial test result reported in the final column). We conclude that, in support of Hypothesis 2.1, men and women believe there to be no performance differences when there is no competition.[25]

<Table 5 about here>

This provides further support for our assumption that there are no descriptive gender stereotypes for this task. When there is (only) social-status ranking, women expect no difference, but men believe that women will underperform relative to men (the binomial test has a significance level of 0.029). This supports Hypothesis 2.2 for men, but not for women. Finally, with only rivalry for resources, men again expect

[25] Ideally, we would prefer to conduct a Bayesian analysis for this hypothesis of no expected gender difference. This would require, however, specifying a hypothesis predicting a specific fraction of women and a specific fraction of men believing that women score better than men. Bayesian analysis could then be used to compare this alternative to our hypothesis 2.1 that these fractions are 0.5. We are unaware of any candidate for such an alternative.

underperformance by women relative to men while women do not. This supports Hypothesis 2.3, but only for men.

Result 2: Male beliefs about gender differences under competition are a possible mechanism underlying the gender differences in performance that we observe.

From this result we conclude that there is a difference between the role of beliefs for men and women, but only under competition. Men think that they will outperform women when submitted to either dimension of competition. Interestingly, these are precisely the cases where we indeed observe that men perform better than women. In this sense, men's beliefs become a self-fulfilling prophecy. On the other hand, it is noteworthy that competition does not appear to activate gender beliefs amongst women. We will discuss this further in the concluding section.

Study 3 – M2: Warmth Stereotype

In Study 3 we consider the second mechanism. Point of departure is that all treatments in Study 1 involving ranking or rivalry have one thing in common. This is that a participant's performance is compared to a group of participants who are performing the cognitive task at the same time. For such contemporaneous comparisons a participant's performance can have a negative impact on others in the group. These consequences stem from either reducing others' ranking (as in treatments with *PR* or *SR*), or reducing others' chances of winning the contest (as with *RfR*), or both. We proposed that competition may activate a *warmth stereotype*, where women are expected to show concerns about such a negative impact. A

warmth stereotype may prescribe showing empathy, or 'caring about others', which may diminish peoples' desires to perform well if doing so harms others.

To identify whether this mechanism is at work we focus on the private ranking environment without rivalry, *nRfR*/*PR*. A priori, there is no reason to believe that warmth would play a different role in distinct dimensions of competition. In all cases, one's success is at the expense of another's failure. We choose to investigate the role of warmth in *nRfR*/*PR*, because this is the environment where an alternative without a negative impact on others is easiest to implement. In this treatment of private ranking but without rivalry, earning a high rank by definition makes other participants be ranked lower than they would have otherwise been. To circumvent this consequence, we created a treatment where this ranking is not vis-à-vis others in the same session. Instead, to determine a private rank, we randomly selected for each participant five other participants from previous *nRfR*/*PR* sessions and anonymously ranked her performance in relation to theirs (these others were not informed about this ranking). In this way, each participant is privately ranked in a group of six peers, but her rank does not affect anyone else's rank. We denote this new treatment as *private historic ranking* (*nRfR/PHR*). For this new treatment, we recruited 65 participants (34 male, 31 female) for three sessions on Feb. 19, 2019 at BLESS in Bologna, Italy, that lasted approximately 60 minutes. Participants earned on average €16.50.

Because there is no impact on others of a good performance, we expect that a warmth stereotype will play no role in this new treatment. Therefore, we expect no gender differences in performance. This yields:

Hypothesis 3

Women and men perform equally well when private ranking does not negatively affect any other participant. That is, there is no gender difference in performance in *nRfR/PHR*.

For a first impression of the results, Figure 5 shows mean performance of men and women in *nRfR/PHR* and compares this to the cases where there is no ranking (*nRfR/nR*) or private ranking within the same session (*nRfR/PR*). The figure suggests that the gender difference in the new treatment is comparable to the benchmark of no competition (*nRfR/nR*), and smaller than when private ranking affects others (*nRfR/PR*). Recall from Table 3 that the gender difference is 1.20 in *nRfR/nR* and 2.71 in *nRfR/PR*. In *nRfR/PHR*, the difference is 0.91 correct summations. This confirms the impression from the figure.

<Figure 5 about here>

Once again, Hypotheses 3 predicts no difference and therefore requires Bayesian analysis. For this purpose, we need to compare a model with no gender difference to a model with such a difference. For the no-difference case, we use *nRfR/nR* (where we observed no gender differences). Alternatively, if a warmth stereotype play no role in the effects of private ranking, then we expect the gender difference in *nRfR/PHR* to be similar to *nRfR/PR*. Thus, we compare a model where *nRfR/PHR* is like *nRfR/nR* to one where *nRfR/PHR* is like *nRfR/PR*. For more details, see Appendix B. The resulting odds ratio is 43 : 1 in favor of the former, indicating that an environment where own performance has no impact on others yields the same result as when there is no competition at all. This provides strong support for Hypothesis 3 and yields:

Result 3: A warmth stereotype activated amongst women under competition is a possible mechanism underlying the gender differences in performance that we observe.

Study 4 – Gender Composition

The goal of this study is to explore whether the process through which competition activates gender beliefs and prescriptive stereotypes and consequently leads to gender differences in performance depends on the *gender composition of the group of competitors*. Our starting point is that a mixed-gender setting is an important scope condition for gender to come into play as a status characteristic. Therefore, in line with status characteristics theory and some empirical findings in behavioral economics, we propose that gendered beliefs may be more strongly activated in mixed-gender environments, where competitors can directly compare themselves to the opposite gender than when all competitors have the same gender. Similarly, we expect that prescriptive stereotypes are more likely to be activated in mixed-gender environments than when one competes only with others of the same gender.

To test these expectations, we organized 16 additional sessions at BLESS (Bologna, Italy), eight of which were all male, and eight with only female participants. These were run between Nov. 28 and Dec. 10, 2019. To test whether the gender composition affects behavior under competition, we ran treatments with either only private ranking (*nRfR/PR*), only social-status ranking (*nRfR/SR*) or rivalry for resources with private ranking (*RfR/PR*). We decided to add information about private ranking to the *RfR* because rivalry always involves some information about private ranking (one learns whether or not one is amongst the top two in a group of

six). As a consequence of adding *PR* to *RfR*, we also included the *nRfR/PR* treatment. A comparison of gender composition effects in *RfR/PR* and *nRfR/PR* then allows us to establish the effect of rivalry alone. Average earnings were €17.10 (excluding the €15 earnings for the C players).

Because we expect that gender beliefs and prescriptive stereotypes are less likely to be activated in same-gender competition than in mixed-gender competition, we predict that women's performance will be negatively affected by mixed-gender competition while men will be positively affected. This gives:

Hypothesis 4

4.1: In the same-gender treatments, women perform better than in the corresponding mixed-gender cases.

4.2 In the same-gender treatments, men perform worse than in the corresponding mixed-gender cases.

For a first impression of the effects of gender composition, Figure 6 shows how men and women respond to same-gender competition.

<Figure 6 about here>

A first thing to notice is that gender differences in performance are more or less the same for mixed- and same-gender groups. In all three same-gender treatments, the gender difference is statistically significant (PtT; *nRfR/PR*: $N = 59$, $p = 0.016$; *nRfR/SR*: $N = 58$, $p = 0.022$; *RfR/PR*: $N = 72$, $p < 0.001$). Thus, gender differences occur in same-gender competition just like we observed in mixed-gender competition. Moreover, the effects of gender composition on the performance of either gender in any treatment appear to be small. To formally test Hypothesis 4.1, we compare for each treatment women's performance in mixed- and same-gender groups. In all cases, the

null of no effect cannot be rejected (PtT; *nRfR/PR*: $N = 61$, $p = 0.328$; *nRfR/SR*: $N = 63$, $p = 0.937$; *RfR/PR*: $N = 72$, $p = 0.656$). For men (Hypothesis 4.2), we also cannot reject the null of no difference in any case (PtT; *nRfR/PR*: $N = 70$, $p = 0.636$; *nRfR/SR*: $N = 64$, $p = 0.270$; *RfR/PR*: $N = 72$, $p = 0.539$). All in all, we conclude that the gender composition does not affect women's or men's performance. As discussed in Section 2, this suggests that cultural gender beliefs and prescriptive stereotypes are also activated when competition is within-gender.

Result 4: The process through which competition activates gender beliefs and prescriptive stereotypes, and leads to the gender differences in performance that we observe does not depend on the gender composition of the group of competitors.

Study 5 – Effort

This study asks whether the differential gender effect of competitive dimensions (via M1 and M2) is caused by gender differences in the effort provided. To answer this question we use the experimental data collected for Study 1, which provides information on men and women's effort levels in the task. The across-treatment pattern we observe for effort should then mirror what we found for performance. This is formalized in hypothesis 5.

Hypothesis 5

Women exert less effort than men when one or both dimension(s) of competition is (are) active, but not when there is no competitive dimension.

We directly measure effort in our experiment by the number of attempted summations. This number captures the visible outcome of effort as opposed to other dimensions, such as innate ability or concentration that are unmeasurable. For a first impression of the results, Figure 7 shows the mean efforts of men and women for each treatment reported in Study 1 (c.f. Figure 4).

<Figure 7 about here>

Eyeballing Figures 4 and 7 suggests that there is little correspondence in the patterns of effort and performance across treatments. Indeed, no clear pattern is obvious in Figure 7. For example, compared to the *nRfR/nR* benchmark of no competition, women make more attempts (on average) in some treatments and fewer in others. The difference with the benchmark varies between –0.53 and +2,24 attempted summations. For women, none of the differences between, on the one hand, effort in the benchmark of no competition and, on the other, effort in any of the competition treatments is statistically significant at the 5% level. Men's effort is also not significantly different than in the benchmark at the 5% level in any of the treatments with competition. As for within-treatment gender differences, the only significant effect at the 5% level that we observe is that men make significantly more effort than women in the benchmark of no competition (PtT, $p = 0.046$, $N = 60$). The benchmark is precisely the scenario where we observed *no* significant gender difference in performance. The big picture is that the strong pattern observed for performance (no significant gender difference in the benchmark of no competition; all differences significant in the competition treatments) is not mirrored in gender differences in effort. We thus reject Hypothesis 5 and conclude that the effort

mechanism cannot explain the performance differences that we observe under competition.

Result 5: The differential gender effect of competition is not caused by gender differences in effort.

GENDER-SPECIFIC RESPONSE TO COMPETITION

The rich data we have collected (altogether more than 750 active participants) allows us to address one final and important question. Given the gender differences that we observe under any competitive dimension, we ask whether this can be attributed to men responding positively to competitive incentives, or to women responding negatively. To optimize statistical power for this analysis we pool data for the two non-competitive environments we have (*nRfR/nR* and *nRfR/PHR*) and also pool all of our competition treatments (*nRfR/PR*, *nRfR/SR*, *RfR/nR*, *RfR/PR*, *RfR/PR*). This gives us 125 observations for the non-competitive setting (61 men, 64 women) and 344 observations with at least one competitive dimension (180 men, 164 women); note that we do not use the same-gender data for this analysis. Figure 8 shows mean performances for these categories.

<Figure 8 about here>

A first thing to observe is that comparisons at this level of aggregation confirm our earlier findings. In the non-competitive case, performance by men and women does not differ significantly (PtT, $N = 125$, $p = 0.122$). When at least one dimension of competition is active, men perform significantly better than women (PtT, $N = 344$, $p < 0.001$). Our goal here, however, is not to compare genders but to compare whether women or men perform differently with competition than without. We observe that

men have on average 1.0 more correct summations when there is competition than when there is not. This difference is marginally significant (PtT, $N = 241$, $p = 0.095$). Women, on the other hand, have 1.4 *fewer* correct summations when there is competition. This difference is highly significant (PtT, $N = 228$, $p = 0.006$). We conclude that competition makes men overperform, but it especially makes women underperform.

DISCUSSION AND CONCLUSION

It has long been recognized that the labor market is a major source of gender inequality. Despite increasing awareness and action, compared to men, women remain in a disadvantageous position. Men receive higher salaries, better jobs, and easier promotions than women do. Such gender inequality is affected by both demand-side and supply-side factors (Gino et al. 2015). Our focus in this study is on the supply side. In particular, we consider a phenomenon that is omnipresent in professional and recreational life, namely competition, and investigate how it differentially affects the behavior of men and women and what the consequences are for gender inequality. The literature on the effects of competitive environments on gender inequality is rich and has been recently dominated by contributions from economics. This literature argues that competition makes men excel in their performance, relative to women (Gneezy et al. 2003). Such studies tend to attribute gender differences in the response to competition –and therefore gender differences in labor market success– to individual preferences and constraints. For example, a mainstream conclusion in the economics literature is that women simply do not like competitive environments (Niederle and Vesterlund 2007, 2011; Niederle 2016). Drawing on status characteristics theory and the stereotype content model, we

argue instead that competition and individual behavior therein are socially constructed. Indeed, the experimental evidence that we present shows that gender differences in performance under competition are endogenous to situational contexts. In this way, insights from sociology and social psychology have proven to be essential for understanding why competition differentially affects men's and women's performance. On the other hand, the sociological literature has to a large extent disregarded the study of competition and its importance for the occurrence of gender inequality. We hope that our study has narrowed the gap between the two disciplines in an attempt to better understand the complex sources of gender inequality.

A first step in our approach is the acknowledgment that competition involves more than just a rivalry for scarce resources. It also creates a status ranking amongst those competing (Schram et al. 2019). Here, we have shown that such a ranking leads to gender inequality in performance that is very similar to the inequality observed under rivalry for resources. Moreover, we show that the two dimensions of competition are substitutes; in any competitive environment where both are active, removing one dimension (for example, by reducing the rivalry) has little effect if the other remains. We also show that the observed gender inequality that is caused by competitive environments is not only driven by men overperforming, but especially by women underperforming. These are remarkable findings that, to the best of our knowledge, have not been established before.

To better understand *why* men and women respond in opposite ways to the dimensions of competition, we systematically studied two explanatory mechanisms that may underlie the observed differences. Our results reveal evidence that both mechanisms are at play and contribute to explaining why competition generates

gender differences. First, cultural gender beliefs about how competition affects men's and women's performance appear to be activated in a competitive environment, but (remarkably) only amongst men. Men believe to be better than women (only) when there is competition, and this belief seems to make men actually perform better. Second, prescriptive stereotypical warmth appears to be activated in women, such that concerns about how a good performance may harm others play a role in women's underperformance when there is competition. In a setting where a good performance did not negatively affect anybody else –while the performer was still ranked vis-à-vis her peers–, women performed as well as men did.

When exploring in more detail how these mechanisms work, we found no evidence that the mixed-gender composition of the group with which one competes moderates the activation of gender beliefs or prescriptive stereotypes and therefore performance. Moreover, our results cannot be attributed to differences in the amount of effort that men and women make when facing a competitive environment.

Because both mechanisms appear to be at work, gender differences in the response to competition cannot simply be attributed to acontextual differences between men and women. Instead, these differences are a product of context, in the way predicted by status characteristics theory and the stereotype content model. Moreover, we contribute to these theories with three important and novel insights. First, we show that gender comes into play as a status characteristic in the social relational context of competition because of the competition *per se*. In our experimental studies, gender beliefs and prescriptive stereotypes are activated by the competitive environment itself and not by the task men and women had to perform. Men do not believe to perform better than women when the same task is

done without competition and women are not affected when their performance does not affect others.

Second, it is remarkable that competition affecting men's beliefs only suffices to yield a gender gap in performance; men's sense of superiority under competition may cause them to excel even if women's beliefs remain unaffected.[26] This difference in men's and women's beliefs might be related to women's growing participation in the labor force: "Perhaps reflecting women's greater labor force involvement, women now describe themselves (but men do not describe them) as significantly more instrumental than did earlier cohorts, narrowing the gender gap in self-descriptions of instrumental competence." (Ridgeway and Correll 2004, p. 527-528).

Third, when considering competition, social-status ranking matters as much as rivalry for resources does. Both activate gender differences in beliefs and prescriptive stereotypes. Remarkably, the two dimensions lead to comparable levels of gender inequality.

To illustrate the implications of our findings, consider again the example of an opening for a professorship. Our results show that, even if men and women are of equal quality *per se*, both the rivalry for the position and the social-status ranking involved will place women at a disadvantage. The reason is that women will underperform because they are expected to fulfil a stereotype and show concern

[26] We measure first-order beliefs in our Study 2. Note, however, that second-order beliefs (which we did not elicit) may also affect performance (Correll et al. 2017). Even though women do not think that they perform worse than men, their performance may be negatively affected by their beliefs that they are expected by others to perform worse than men.

about what an appointment would do to other candidates, while men will excel because they believe that the competitive environment is to their advantage.

Our research is based on a series of laboratory experiments, which naturally raises questions about the external validity of our results. What do these findings tell us about the world outside of the laboratory? Much of the previous research has shown how the results from experiments relate to what happens under competition in natural environments, in particular in education and at the workplace[27]. This external validity is important, but it is equally important to realize that the choice of laboratory experiments as our method is founded in their internal validity. For the questions we ask, laboratory control is important. It allows us to isolate the two dimensions of competition and directly measure the causal effect each has on performance. It also allows us to directly measure performance, effort, and beliefs. Finally, by introducing simple changes to the design of study 1, while keeping all other aspects of the design constant, we were able to systematically investigate the effects of the explanatory mechanisms. None of these analyses would be possible with observational field data. The possibilities that laboratory control offers thus make it the most suitable method for our purposes. As an alternative, one could consider collecting data in field experiments instead. Indeed, some of the questions we ask could also be addressed by properly designed field experiments. Others, however, require a level of control that would override possible advantages of experimentation in the field. In particular, studying competition while isolating

[27] For example, laboratory competitiveness explains why women avoid jobs with competitive compensation regimes (Flory, Leibbrandt, and List 2015) and predicts student participation in a competitive university entry exam (Zhang 2013). It also predicts American students' expectations about future salaries (Reuben, Wiswall, and Zafar 2017) and explains gender differences in academic career choices (Buser, Niederle, and Hessel 2014).

status ranking from rivalry for resources seems hard to realize with the diminished control that is inherent to field designs. In our view, the internal validity offered by the laboratory outweighs at this stage possible considerations of higher external validity in the field.

The fact that our conclusions rest on laboratory data provides a solid first step to better understand what could possibly be going on in the world outside. Future research, either through field experiments or observational data, can build on our findings obtained under laboratory control and explore further their replication as well as implications to the world outside of the laboratory. The fact that various studies have indeed found that behavior under competition in the laboratory is highly informative for actual educational and labor environments gives us confidence that this future research will find our results to be very useful to understand gender inequalities in the world at large.

It is vital for decision-makers in all sorts of organizations to be aware of gender differences in the response to competition if they wish to provide an equitable work environment that fosters the organization as a whole. We cautiously put forward some suggested policies aimed at reducing gender inequality. First, when it comes to mitigating the effects of a competitive setting, both the rivalry for resources and the social-status ranking dimensions of competition should be considered; addressing only one does not reduce gender inequality. Second, updating information and increasing awareness on men and women's true abilities may help to reduce the gender gap. Such information should target both men and women and emphasize that if women can do as well as men without competition, they are capable of doing equally well with competition. A final and perhaps more radical suggestion to reduce gender inequality is that organizations should

reconsider their competitive models and create an environment where one's hiring or promotion is not necessarily determined by the assessment of one's relative performance to others but by the performance *per se* (based on a set of pre-determined criteria), irrespective of how others perform.

References

Azmat, Ghazala, and Barbara Petrongolo. 2014. "Gender and labor market: What Have We Learned from Field and Lab Experiments." *Labour Economics* 30: 32-40.

Ball, Sheryl B., Catherine C. Eckel, Philip J. Grossman, and William R. Zame. 2001. "Status in Markets." *Quarterly Journal of Economics* 116(1): 161-188.

Berger, James O. 2013. *Statistical decision theory and Bayesian analysis*. New York: Springer Science & Business Media.

Berger, Joseph, Thomas L. Conner, and M. Hamit Fisek. 1974. *Expectation States Theory: A Theoretical Research Program.* Cambridge, MA: Winthrop Publishers.

Berger, Joseph, David G. Wagner, and Morris Jr. Zelditch. 1985. "Expectation States Theory: Review and Assessment." Pp. 1-72 in *Status, Rewards, and Influence: How Expectations Organize Behavior*, edited by Joseph Berger, and Morris Jr. Zelditch. San Francisco: Jossey-Bass.

Biernat, Monica, and Diane Kobrynowicz. 1997. "Gender-and Race-Based Standards of Competence: Lower Minimum Standards but Higher Ability Standards for Devalued Groups." *Journal of Personality and Social Psychology* 72(3): 544.

Brandts, Jordi, Klarita Gërxhani, and Arthur Schram. 2020. "Are There Gender Differences in Status-Ranking Aversion?" *Journal of Behavioral and Experimental Economics* 84: 101485.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107(11): 3288-3319.

Buser, Thomas, Alexander W. Cappelen, and Bertil Tungodden. 2021. "Fairness and Willingness to Compete." *Discussion paper SAM 08/2021, Norwegian School of Economics.*

Buser Thomas, Muriel Niederle, and Hessel Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics* 129(3): 1409–1447.

Castilla, Emilio J., and Stephen Benard. 2010. "The Paradox of Meritocracy in Organizations." *Administrative Science Quarterly* 55(4): 543-576.

Ceci, Stephen J., and Wendy M. Williams. 2010. "Sex Differences in Math-Intensive Fields." *Current Directions in Psychological Science* 19(5): 275–279.

Coffman, Katherine B., Christine L. Exley, and Muriel Niederle. 2021. "The Role of Beliefs in Driving Gender Discrimination." *Management Science* 67(6): 3551-3569.

Connor, Rachel A., and Susan T. Fiske. 2018. "Warmth and Competence: A Feminist Look at Power and Negotiation." Pp. 321-342 in *APA Handbook of the Psychology of Women: History, Theory, and Battlegrounds, Vol. 1*, edited by Cheryl B. Travis, Jacquelyn W. White, Alexandra Rutherford, Wendi S. Williams, Sarah L. Cook, Karen F. Whyche. American Psychological Association.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessments." *American Journal of Sociology* 106(6): 1691-1730.

Correll, Shelley J. 2004. "Constraints into preferences: Gender, status, and emerging career aspirations." *American Sociological Review* 69(1): 93-113.

Correll, Shelley J., and Stephen Benard. 2006. "Biased Estimators? Comparing Status and Statistical Theories of Gender Discrimination." *Advances in Group Processes* 23: 89-116.

Correll, Shelley J., and Cecilia L. Ridgeway. 2006. "Expectation States Theory." Pp. 29-51 in *Handbook of Social Psychology*, edited by John Delamater. Boston: Springer.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. "It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage." *American Sociological Review* 82(2): 297-327.

Deaux, Kay, and Marianne LaFrance. 1998. "Gender." Pp. 788-828 in *The Handbook of Social Psychology*, edited by Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey. New York: Oxford University Press.

Dickerson, Sally, and Margaret Kemeny. 2004. "Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research." *Psychology Bulletin* 130(3): 355-391.

DiMaggio, Paul. 1997. "Culture and Cognition." A*nnual Review of Sociology* 23(1): 263-287.

Ebert, Irena D., Melanie C. Steffens, and Alexandra Kroth. 2014. "Warm, but Maybe Not So Competent?—Contemporary Implicit Stereotypes of Women and Men in Germany." *Sex roles* 70(9-10): 359-375.

Erickson, Kristan G. 1998. "The Impact of Cultural Status Beliefs on Individual Task Performance in Evaluative Settings: A New Direction in Expectation States Research". Ph.D. dissertation. Stanford University.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow from Perceived Status and Competition." *Journal of Personality and Social Psychology* 82(6): 878-902.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Gender Differences in Job-entry Decisions." *Review of Economic Studies* 82(1): 122–155.

Foschi, Martha. 2000. "Double Standards for Competence: Theory and Research." *Annual Review of Sociology* 26(1): 21-42.

Foschi, Martha, Larissa Lai, and Kirsten Sigerson. 1994. "Gender and Double Standards in the Assessment of Job Applicants." *Social Psychology Quarterly* 57(4): 326-339.

Frank, Robert H. 2004. *What Price the Moral High Ground?*. NJ: Princeton University Press.

Gërxhani, Klarita. 2020. "Status Ranking and Gender Inequality: A Cross-Country Experimental Comparison." *Research in Social Stratification and Mobility* 65: 100474.

Gino, Francesca, Caroline A. Wilmuth, and Alison W. Brooks. 2015. "Compared to Men, Women View Professional Advancement as Equally Attainable, but Less Desirable." *Proceedings of the National Academy of Sciences* 112(40): 12354-12359.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics* 118(3): 1049–1074.

Heilman, Madeline E., and Tyler G. Okimoto. 2007. "Why Are Women Penalized for Success at Male Tasks?: The Implied Communality Deficit." *Journal of Applied Psychology* 92(1): 81–92.

Heilman, Madeline E., and Elizabeth J. Parks-Stamm. 2007. "Gender Stereotypes in the Workplace: Obstacles to Women's Career Progress." Pp. 44-77 in *Social Psychology of Gender (Advances in Group Processes, Vol. 24)*, edited by Shelley J. Correll. Bingley: Emerald Group Publishing Limited.

Heilman, Madeline E., Aaron S. Wallen, Daniella Fuchs, and Melinda M. Tamkins. 2004. "Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks." *Journal of Applied Psychology* 89(3): 416-427.

Hobson, Barbara. 1990. "No Exit, No Voice: Women's Economic Dependency and the Welfare State." *Acta Sociologica* 33(3): 235-250.

Kazdin, Alan E. 2007. "Mediators and Mechanisms of Change in Psychotherapy Research." *The Annual Review of Clinical Psychology* 3:1–27.

King, Molly M, Carl T. Bergstrom, Shelley J. Correll, Jennifer Jacquet, Jevin D. West. 2017. "Men Set Their Own Cites High: Gender and Self-Citation across Fields and over Time." *Socius* 3. doi:10.1177/2378023117738903

Koedel, Cory, and Eric Tyhurst. 2012. "Math Skills and Labor-Market Outcomes: Evidence from a Resume-Based Field Experiment." *Economics of Education Review* 31(1): 131-140.

Kuhnen, Camelia M., and Agnieszka Tymula. 2012. "Feedback, Self-Esteem and Performance in Organizations." *Management Science* 58(1), 94-113.

Lovaglia, Michael J., Jeffrey W. Lucas, and Shane R. Thye. 1998. "Status Processes and Mental Ability Test Scores." *American Journal of Sociology* 104(1): 195-228.

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56 (3): 799-866.

Niederle, Muriel. 2016. "Gender." Pp. 481-553 in *Handbook of Experimental Economics*, second edition edited by John Kagel, and Alvin E. Roth. Princeton: Princeton University Press.

Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete too much?" *Quarterly Journal of Economics* 122(3): 1067-1101.

Niederle, Muriel, and Lise Vesterlund. 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24(2): 129-44.

Niederle, Muriel and Lise Vesterlund. 2011. "Gender and Competition." *Annual Review of Economics* 31(1): 601-630.

Penner, Andrew M., and Robb Willer. 2019. "Men's Overpersistence and the Gender Gap in Science and Mathematics." *Socius* 5: 2378023118821836.

Prentice, Deborah A., and Erica Carranza. 2002. "What Women and Men Should be, Shouldn't be, are Allowed to be, and don't Have to be: The Contents of Prescriptive Gender Stereotypes." *Psychology of Women Quarterly* 26(4): 269–281.

Quadlin, Natasha. 2018. "The Mark of a Woman's Record: Gender and Academic Performance in Hiring." *American Sociological Review* 83(2): 331-360.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25: 111-163.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender." *Economic Journal* 127(604): 2153–2186.

Ridgeway, Cecilia L. 2014. "Why Status Matters for Inequality." *American Sociological Review* 79(1): 1-16.

Ridgeway, Cecilia L., and Shelley J. Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender & Society* 18(4): 510-531.

Ridgeway, Cecilia L., and Lynn Smith-Lovin. 1999. "The Gender System and Interaction." *Annual Review of Sociology* 25(1): 191-216.

Rivera, Lauren A. 2017. "When Two Bodies are (not) a Problem: Gender and Relationship Status Discrimination in Academic Hiring." *American Sociological Review* 82(6): 1111-1138.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." *Journal of Political Economy* 129(1): 101-147.

Schram, Arthur, Jordi Brandts, and Klarita Gërxhani. 2019. "Status Ranking: A Hidden Channel to Gender Inequality under Competition." *Experimental Economics* 22(2): 396-418.

Shurchkov, Olga. 2012. "Under Pressure: Gender Differences in Output Quality and Quantity Under Competition and Time Constraints." *Journal of the European Economic Association* 10(5): 1189–1213.

Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology* 35(1): 4-28.

Steele, Claude M. 1997. “A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance.” *American Psychologist* 52(6): 613–629.

Stryker, Sheldon, and Kevin D. Vryan. 2006. “The Symbolic Interactionist Frame.” Pp. 3-28 in *Handbook of Social Psychology*, edited by John Delamater. Boston: Springer.

Stuijfzand, Suzannah, Minet De Wied, Maaike Kempes, Jolien Van de Graaff, Susan Branje, and Wim Meeus. 2016. “Gender Differences in Empathic Sadness Towards Persons of the Same-Versus Other-Sex During Adolescence.” *Sex roles* 75(9): 434-446.

Van Dolder, Dennie, Martijn Van Den Assem, and Thomas Buser. 2020. “Gender and Willingness to Compete for High Stakes.” *Tinbergen Institute Discussion Paper (No. TI 2020-011/I).*

Wagner, David G., and Joseph Berger. 1997. “Gender and Interpersonal Task Behaviors: Status Expectation Accounts.” *Sociological Perspectives* 40(1): 1-32.

Weber, Matthias, and Arthur Schram. 2017. “The Non-Equivalence of Labor Market Taxes: A Real-Effort Experiment.” *The Economic Journal* 127(604): 2187-2215.

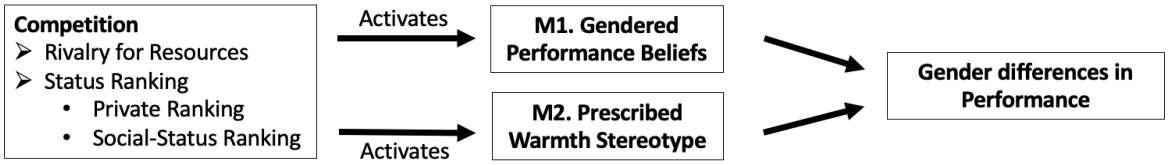
Willer, Robb, Christabel L. Rogalin, Bridget Conlon, and Michael T. Wojnowicz. 2013. “Overdoing Gender: A Test of the Masculine Overcompensation Thesis.” *American Journal of Sociology* 118(4): 980-1022.

Willer, Robb, Christopher Wimer, and Lindsay A. Owens. 2015. “What Drives the Gender Gap in Charitable Giving? Lower Empathy Leads Men to Give Less to Poverty Relief.” *Social Science Research* 52: 83-98.

Zhang, Y. Jane. 2013. “Can Experimental Economics Explain Competitive Behavior Outside the Lab?” Working paper, Hong Kong University of Science and Technology.

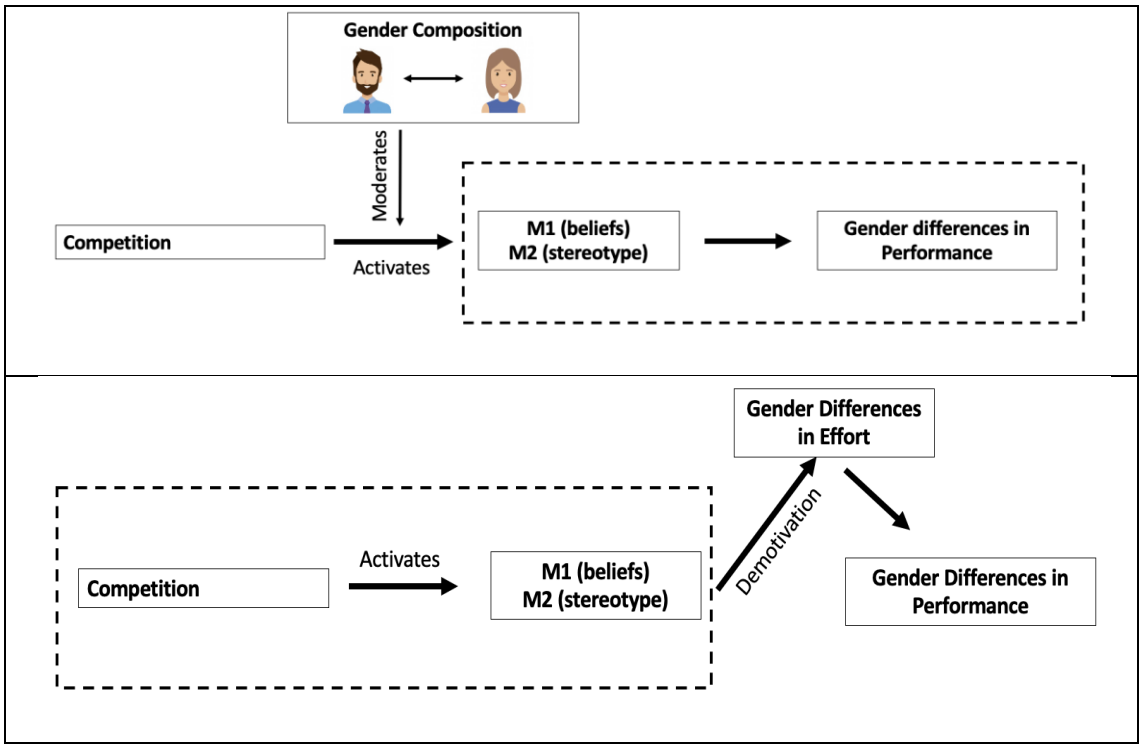
List of figures

Figure 1: Principal Mechanisms



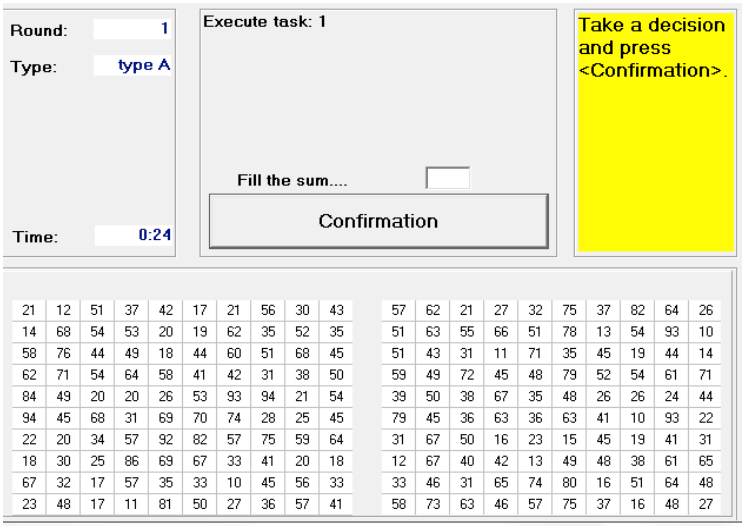
Notes. The dimensions of competition cause gender differences in performance via mechanism M1 if the dimensions activate gender differences in performance beliefs, which in turn yield gender differences in performance. Mechanism M2 leads to such differences if the dimensions of competition activate stereotypes where women are prescribed to show warmth towards others in the competition.

Figure 2: Fine-tuning the Principal Mechanisms



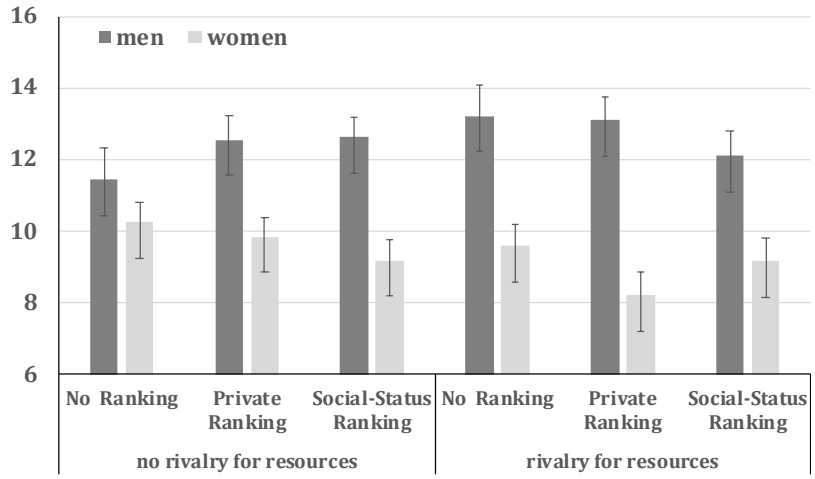
Notes. Upper panel: the gender composition of the competitors determines the extent to which competition activates gendered performance beliefs and a prescribed warmth stereotype; these beliefs and stereotypes yield gender differences in performance. Because we cannot separately measure whether gender composition moderates the activation of gender beliefs and stereotypes under competition, but argue that it does, we jointly depict the effects of gender composition on the mechanisms and performance. Lower panel: competition activates gendered beliefs and a prescribed warmth stereotype; these two mechanisms demotivate individuals and yield gender differences in effort that result in gender differences in performance. Because we assume competition affects effort indirectly via M1 and M2, these are jointly depicted.

Figure 3: Screenshot Part 1



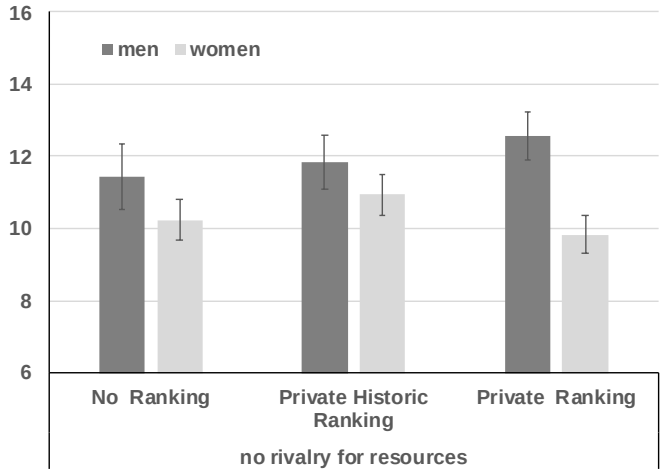
Notes. As explained in Schram et al. (2019), the instructions inform participants that numbers were 'randomly generated'. Drawing from a uniform distribution would give a high probability of very high sums. To avoid this, for each cell, we first drew a random number between 40 and 99, say X. Then, we drew a random number (uniformly) between 10 and X. This gives a far lower probability of high numbers (the chance of a number being 75 or more is approximately 0.06).

Figure 4: Performance and Competition



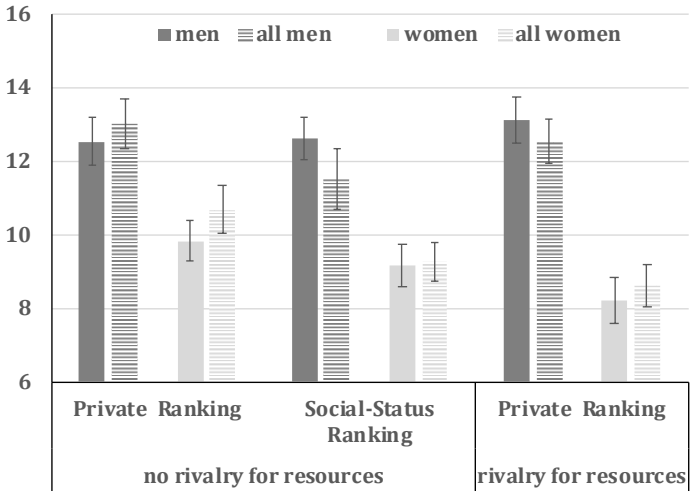
Notes. Bars show the mean number of correct summations for men and women in each treatment. Error bars show standard errors.

Figure 5: Performance and Warmth Stereotype



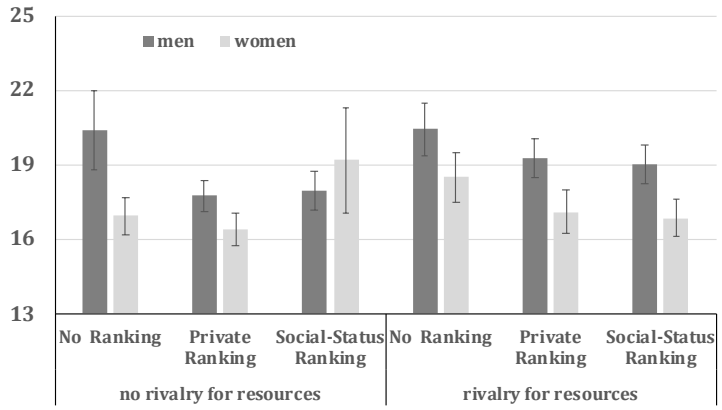
Notes. Bars show the mean number of correct summations for men and women in each treatment. Error bars show standard errors.

Figure 6: Performance and Gender Composition



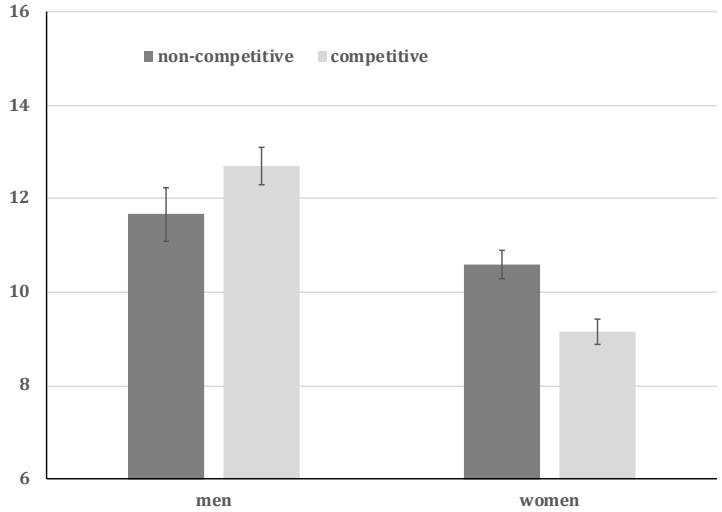
Notes. Bars show the mean number of correct summations for men and women. Neighboring bars show the mixed-gender and same-gender cases. In each treatment there are four bars: men in mixed-gender competition (dark-solid gray), men in all-men competition (dark-striped gray), women in mixed-gender competition (light-solid gray), women in all-women competition (light-striped gray). Error bars show standard errors.

Figure 7: Effort and Competition



Notes. Bars show the mean number of attempted summations for men and women in each treatment. Error bars show standard errors.

Figure 8: The Effects of Competition per Gender



Notes. Bars show the mean number of correct summations. Error bars show standard errors.

List of tables

Table 1: Treatments

Treatment	Player type	Information	Payoff
No ranking, no rivalry for resources	B	none	piece rate
Private ranking, no rivalry for resources	B	rank	piece rate
Social-status ranking, no rivalry for resources	A	rank	piece rate
No ranking, rivalry for resources	B	none	tournament
Private ranking, rivalry for resources	B	rank	tournament
Social-status ranking, rivalry for resources	A	rank	tournament

Notes. *A* players report to the *C* player, *B* players do not. 'Information' denotes whether a participant is told her rank within the group. In 'piece-rate' payoffs, every player is rewarded for the own score. In 'tournament' only the top two performers in a group are rewarded.

Table 2: Numbers of Observations

		male	female
Non Rivalry (nRfR)	No Status Ranking (nR)	27	33
	Private Ranking (PR)	40	32
	Social-Status Ranking (SR)	35	37(3)
Rivalry (RfR)	No Status Ranking (nR)	30	30
	Private Ranking (PR)	36	36
	Social-Status Ranking (SR)	39	33(1)
Total		207	201

Notes. Cells show the number of observations by treatment. Numbers in parentheses show the number of outliers in a cell. Outliers are defined as participants who repeatedly enter summation larger than 200 (which is impossible with the maximum in each matrix being 100). Outliers were excluded from the analysis. Our results do not change if we include these outliers.

Table 3: Gender Differences

	No Rivalry for Resources (nRfR)			Rivalry for Resources (RfR)		
	nR	PR	SR	nR	PR	SR
Difference	1.20	2.71	3.45	3.63	4.89	2.94
N	60	72	69	60	72	71
p-value	0.256	0.004	<0.001	0.001	<0.001	0.004

Notes. *nR*: No Status Ranking; *PR*: Private Ranking; *SR*: Social-Status Ranking. 'Difference' shows the difference between men and women in mean performance (measured as the number of correct summations), with a positive number indicating that men perform better. *N* is the number of observations, and the *p*-value is the result of a PtT testing against the null of no gender difference in mean performance.

Table 4: Substitutes or Complements?

	RfR/PR to	RfR/PR to	RfR/SR to	RfR/SR to
Baseline comparison	nRfR/PR	RfR/nR	nRfR/SR	RfR/nR
Benchmark comparison	RfR/nR	nRfR/PR	RfR/nR	nRfR/SR
Odds ratio	1 : 2.3	1.2 : 1	28.3 : 1	24.8 : 1

Notes. Odds ratios report the likelihood that a model that assumes that the two dimensions are substitutes is correct divided by the likelihood that a model that assumes they are complements is correct. *RfR* = rivalry for resources; *PR* = private ranking; *SR* = social-status ranking. The 'baseline comparison' x to y refers to a comparison of (1) gender differences where dimensions x and y are both included to (2) the gender differences when only dimension y holds.

Table 5: Beliefs

		fraction	N	p-value
No Competition	women	0.40	48	0.193
	men	0.42	48	0.312
Social-Status Ranking	women	0.40	48	0.193
	men	0.33	48	0.029
Rivalry for Resources	women	0.50	48	1.000
	men	0.33	48	0.029

Notes. The column 'fraction' shows the fraction of participants (per gender) who think that women perform better than men. The *p*-values in the last column refer to a binomial test that the fraction concerned is equal to 0.5.