Research in Social Stratification and Mobility xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Research in Social Stratification and Mobility

journal homepage: www.elsevier.com/locate/rssm

Status ranking and gender inequality: A cross-country experimental comparison

Klarita Gërxhani

Department of Political and Social Sciences, European University Institute, Via dei Roccettini 9, 50014, San Domenico di Fiesole, FI, Italy

ARTICLE INFO

Keywords:
Status ranking
Gender
Experiments
Culture

ABSTRACT

This article examines the (differential) effects of status ranking on men's and women's performance. It first recognizes that status ranking might be implicit or explicit. Then, it theoretically studies and predicts the gender effects of both types of status ranking and how these effects might vary with culture. Finally, an empirical analysis is presented based on conducting the same experimental design in three culturally different countries, i.e., Italy, the Netherlands and Spain. The experimental results provide evidence that both the type of status ranking and culture matter. When status ranking is explicit, strong gender differences in performance are observed. In more masculine and competitive environments like the Italian and the Spanish, women perform significantly worse than men. Importantly however, cultural beliefs about gender are not sufficient to drive gender differences when these beliefs are the basis for implicitly inferred status ranking among men and women. It appears that more is needed for gender inequality to kick in than an implicit inference from status characteristics.

1. Introduction

Social status, typically defined as a cultural construct based on widely shared evaluations of where different 'types' of people should stand in a social hierarchy (Berger, Cohen, & Zelditch, 1972), is an important factor in social stratification. "People care about status as much as they care about money, and status is an important element of social inequality" (Ridgeway, 2014: 13). The importance of social status has been recognized in the classic socio-economic literature (Merton, 1968; Veblen, 1931; Weber, 1968 [1922; Weber, 1968 [1922]), and both theoretical and empirical research on social status has revived in recent years, across social psychology, economics and sociology (e.g. Anderson, Willer, Kilduff, & Brown, 2012; Correll & Ridgeway, 2003; Heffetz & Frank, 2008).

Social status involves the social recognition of one's standing in a hierarchy (Ball, Eckel, Grossman, & Zame, 2001), which in turn can be determined by one's economic resources and others' evaluation of such resources. In search of this recognition, people may seek to increase their status in trying to 'Keep up with the Joneses' (Veblen, 1931), which can lead to more effort and higher achievement (Parsons, 1970). Alternatively, people may 'Give up on the Joneses' altogether because of feelings of status anxiety and threats of losing social esteem (Dickerson & Kemeny, 2004; Paskov, Gërxhani, & van de Werfhorst, 2017). Hence, anticipating that one's social ranking will be known (i.e. salient) seems to affect one's performance positively or negatively (De Botton, 2004; Wilkinson & Picket, 2010).

Until very recently, little was known about whether the effect of status ranking on performance was different for men and women (Schram, Brandts, & Gërxhani, 2019). There is some reason to believe that such gender differences exist. Numerous (experimental) economics studies on competition have shown that (compared to men) women underperform in and avoid competitive environments (Gneezy, Niederle, & Rustichini, 2003; Niederle & Vesterlund, 2007). The focus of these studies has been on one particular dimension of competition, namely the rivalry for resources (Stigler, 1987).

Schram et al. (2019) argue, however, that a second, equally important dimension of competition has been overlooked, namely the ranking of relative performance as a result of competition. In many settings competition not only allocates resources, it also provides a social ranking of the competitors. For example, in competition for a job the successful candidate may be seen as having been ranked higher than his or her competitors. When such a performance ranking is socially recognized then it leads to a social-status ranking, as defined by Ball et al. (2001). In a competitive environment, the rivalry and status di-

E-mail address: Klarita.Gerxhani@eui.eu.

https://doi.org/10.1016/j.rssm.2020.100474
Received 14 March 2019; Received in revised form 19 November 2019; Accepted 7 January 2020

mension might go hand in hand, or one of the two might be more salient than the other.[1]

Schram et al. (2019) introduce an experimental design that isolates the social ranking dimension of competition from the rivalry for resources. There, the main finding is that "men make more attempts and increase their performance in anticipation of status ranking. Women, on the other hand, make fewer attempts and perform more poorly when they know they will be compared to others. This results in a large and statistically highly significant gender gap." (p. 14). When considering this result, it is important to distinguish between 'implicit' and 'explicit' status ranking. In the former, the ranking of one's competence is inferred from a status characteristic like gender. Implicit status ranking is the focus of status characteristics theory (Berger, Fisek, Norman, & Zelditch, 1977; Ridgeway, 1991). In contrast, Schram et al. (2019) test the causal effect of 'explicit' status ranking, where one's own competence relative to others' is explicitly given.

This study starts by recognizing that there are many environments where status ranking may be explicit (think of schools, workplaces, sport teams, etc.). I then formally distinguish between the two types of status ranking and address the question of whether implicit and explicit status ranking are complements or substitutes in their effect on gender performances. This article thus offers its first main contribution to the literature by theoretically recognizing and empirically isolating the causal effect of both implicit and explicit status rankings on gender performances.

Moreover, it does so in different cultural settings. The results of Schram et al. (2019) were obtained in Barcelona, Spain. This raises the question of whether they are context-driven, bound by the Spanish case. If social status is a cultural construct, one may not find similar results across different cultures. Status characteristics theory argues that gender is a status characteristic if it conveys widely held *cultural beliefs* in attributing more ability to one category of the characteristic (e.g. men) compared to the other (women). Yet, the cultural dimension is often assumed. In contrast, this study explicitly tests the cultural dimension. It does so by looking at the effects of implicit and explicit status ranking on gender performances in different cultural settings. The investigation of the role of cultural beliefs in the effects of status ranking is the second main contribution of this study. For this purpose, I loosely define culture as a system of common beliefs and values that govern individuals' interactions in a particular context. Culture and its impact on the organization of a society is central to sociological research (Bourdieu, 1984; DiMaggio, 1994). Because cultural beliefs play a central role in status characteristics theory, there is a priori little reason to expect that gender differences in performance due to status ranking will be the same across cultures. The results obtained in Barcelona may not be replicated in other cultures.

To study this, I compare the results of the same experiment conducted in three culturally different countries, i.e. Spain, the Netherlands and Italy. In this way, this article offers the first experimental cross-country comparison of gender differences in the performance consequences of social-status ranking. Traditionally, comparative studies have been based on observational field data. Because many aspects differ between any two or more countries, it is often difficult with such data to attribute a difference in the subject of interest to any specific difference between the countries. As advocated by Falk and Heckman (2009), conducting cross-country laboratory experiments provides experimental control over the environment that allows one to correct for between-country differences on external constraints. This allows the researcher to isolate causal relations under truly ceteris paribus conditions. The laboratory rules of the game (institutions) in this study are held constant across countries. This enables an investigation of how - above and beyond these rules - cross-country cultural differences affect the way in which implicit and explicit status ranking affect gender performances.

If cultural differences matter, then the selection of countries to study may be important. As a first country, the Barcelona experiment was replicated in Amsterdam, the Netherlands. This allows me to test for cultural differences in a country rather opposite to Spain when it comes to both gender policies and gender beliefs (Minguez, 2010). For the second replication, Bologna, Italy was chosen. This allows for a stress test of the Spanish findings in a similar cultural setting. Spain, however, has recently shown progressive movements towards more gender equality.[2] According to the European Institute for Gender Equality (2017), the gender equality index in 2017 for Spain was 68.3 out of 100 (slightly above the EU-28 score of 66.2), for the Netherlands it was 72.9 and for Italy 62.1. These scores are in line with Hofstede, Hofstede, and Minkov's (2010) classification, where Italy and Spain are culturally closer to each other, than either is to the Netherlands. Yet, some differences between Italy and Spain appear when it comes to gender-related cultural expectations. In the following section, I will use Hofstede et al. (2010) renowned model of national culture to predict cross-country differences because of the specificities it offers with respect to country-aggregated gender beliefs and competition.

2. Theoretical discussion and hypotheses

2.1. Status-ranking

According to the *social self-preservation theory* in social psychology, the social self reflects one's social value, esteem, and status and is largely based on others' perceptions of one's worth (de Waal, 1989; Dickerson & Kemeny, 2004; Gilbert, 1997). Hence, the preservation of the social self depends on others' social assessment, which in turn can lead to social comparisons. This implies a comparison between oneself and the rest of the (reference) group, which effectively determine one's status as defined above. The social self-preservation theory further argues that such a comparison may evoke threats to one's self-image because of being evaluated in comparison to others. In the presence of such social evaluative threats, one may feel embarrassed or ashamed and ultimately experience higher levels of anxiety, which in turn are expected to affect one's performance. There is substantial empirical evidence of such a negative effect of social comparison on performance (see Dickerson & Kemeny, 2004 for an overview). Nevertheless, the empirical findings do not report any gender differences in how men and women respond to status threats. This leads us to expect that the anticipation of status ranking will affect the performance of both men and women negatively, but it does not predict differential effects for men and women.

According to the *status characteristics theory* in sociology (Berger et al., 1977; Ridgeway, 1991), however, gender differences can be expected in the effect of status on performance. This theory argues that in uncertain environments where one's worth or ability is unknown, an observable general characteristic that differentiates individuals –like gender or race– is used to infer performance expectations. "Gender is commonly described as a diffuse status characteristic, meaning that widely shared cultural beliefs about gender include expectations that men are diffusely more competent or capable at most things [compared to women]" (Correll, 2004: 97). The psychology literature argues that

[1] For situations where rivalry for resources is less important and status ranking is salient, think for example of positions in universities, the military, the judiciary, and the churches. In these professions people in high positions typically enjoy high status, while payoff differences across different ranks are often not large. Thus, they do not necessarily 'compete' for resources (i.e., a job or higher salary), but for higher status. For example, a tenured Harvard professor might not compete with colleagues from other universities for a position, but they do compete for academic status.

[2] https://elpais.com/elpais/2018/03/08/inenglish/1520500265_433354.html.

such cultural beliefs about gender-specific performance can lead to stereotype threats, which in turn may lead to evaluation anxiety, affecting men's and women's performance differently (Fiske, 1998; Steele, 1997).

Originally, in line with expectation states theory (Webster & Foschi, 1988), status characteristics theory argued that status ranking emerges only in groups that have a collective goal or task. This is because it is the restriction of successfully achieving the collective goal that 'force' group members to rely on a status characteristic to infer performance expectation and hence status ranking within a group. New developments within the theory, however, argue that the impact of gender as a status characteristic goes beyond collective group orientation (Correll, 2001). As extensively discussed in Correll and Ridgeway (2003), a relative comparison of self to others, even when undertaking an individual (and not a collective) task, is all that is needed for the theory to hold. In line with the social self-preservation theory and the social evaluative threat, discussed above, the main argument here is that "individual evaluative tasks can provide the pressure to make relative assessments of competence in situations where actors know they will receive a socially important and socially valid performance evaluation." (Correll & Ridgeway, 2003: 47). In an uncertain environment, where others' and own competence are fairly unknown, the expectation of implicitly being compared to others will also 'force' people to rank themselves in relation to others. As long as the 'true' ability remains ambiguous, salient status characteristics 'will give people a hand' in simplifying the complex situation and inferring performance expectations according to cultural beliefs they hold about men's and women's competence related to the individual task.

There is by now substantial empirical evidence that in mixed-gender situations and especially in what are perceived to be male-oriented tasks (e.g., mathematical tasks), men are culturally believed to be more competent than women; in such tasks, then, being a man will infer a higher status than being a woman. As a result, men will assess their abilities higher than women, which will lead to gender differences in individual performance (Correll, 2004; Ridgeway, 2001). Following this line of reasoning, the expectation is that in an uncertain environment where competence is unknown, the 'inferred' status ranking will make women underperform compared to men.

Importantly, status characteristics theory is based on implicit (i.e., inferred) status ranking. In previous studies, there is no explicit ranking or comparison with others. A question arises on what will happen if the own versus others' competence is made known, that is, explicitly given. In this case, status ranking is no longer just implicit. Recent evidence shows that women perform worse than men in anticipation of explicit status ranking (Schram et al., 2019). Thus, the expectation formulated above that women will underperform compared to men in environments of implicit status ranking, appears to carry over to situations where the ranking of relative performance is made explicit. In short, women are expected to perform worse than men whether status ranking is implicit or explicit. This assessment will form the basis for the formal hypotheses, below.

2.2. Cultural differences

As argued above, cultural beliefs are central to status-related theories. For this reason, this article takes a closer look at the possibility of cultural variation in gender differences in the effects of status ranking on performance. It does so by comparing three culturally different countries, namely Italy, the Netherlands and Spain. Hofstede's et al. (2010) model of national culture is used to broadly specify the core cultural dimensions characterizing each of the three countries.

In Hofstede's setup, there are six dimensions:

(1) Power distance; the (in)equality in the distribution of power in a society;
(2) Individualism; the extent to which individuals look after themselves and their families only;
(3) Masculinity-Femininity; the extent to which a society is characterized by competitive and achievement-oriented values versus caring for others and the quality of life;
(4) Uncertainty avoidance; the degree to which individuals feel threatened by unknown situations and would prefer to avoid these;
(5) Long term orientation; the extent to which a society is more present and future oriented than holding on to the past;
(6) Indulgence; the degree to which individuals are willing to fulfil their impulses and desires, and be positive and optimistic in life.

Originally (1968–1972), Hofstede formulated these dimensions by surveying employees of the IBM corporation on various questions about cultural values. The number of countries, questions and professions have increased since, with the latest scores reported in Hofstede et al. (2010), where 72 countries and regions are included. Using factor analysis, Hofstede and his colleagues grouped the survey questions into these six dimensions, and calculated mean scores per dimension and country. The score scale runs from 0 to 100.

Fig. 1 below displays a comparison between Italy, the Netherlands, and Spain along these six dimensions.

It appears that the Netherlands has the lowest scores on the power distance, masculinity, and uncertainty avoidance dimensions and the highest scores on individualism, long term orientation and indulgence. In other words, the Netherlands is a relatively equal society, where individuals feel independent, prefer compromise over competition, and do not feel very anxious about the future. The opposite seems to be the case for Italy and Spain, although some differences do exist between

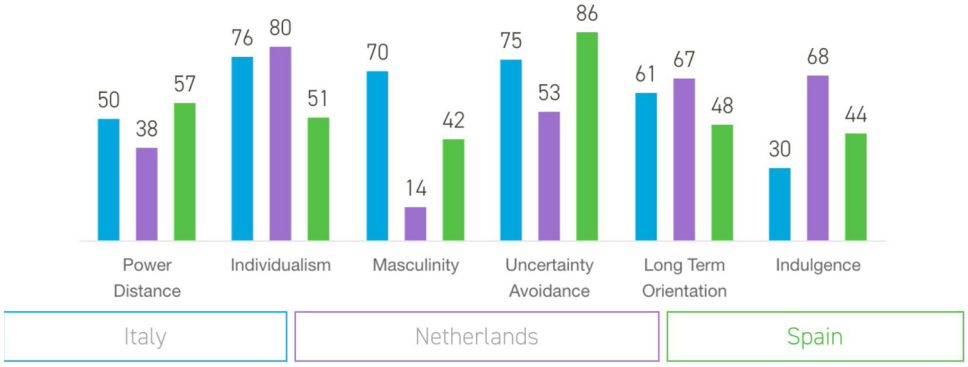

Fig. 1. Country comparison along 6 cultural dimensions.
Notes. Bars indicate the mean scores per dimension and country on various survey questions on cultural values.
Source: Hofstede Insights - Country Comparison. URL: https://www.hofstede-insights.com/country-comparison/.

these two Mediterranean countries. In particular, Italian society seems to be more individualistic and masculine compared to Spain.

For cultural variation on gender differences in the effects of status ranking on performance, the Masculinity-Femininity dimension appears to be most relevant. Recall that a higher score on the Masculinity dimension implies a stronger inclination towards competition and achievement. In countries that score highly on masculinity (and, therefore, lowly on femininity), competitiveness plays an important role. I conjecture that this makes the social evaluative threat in such societies stronger than in countries that tend more towards femininity. Focusing specifically on this dimension, Fig. 1 shows a very sharp difference between countries, with the Netherlands being the least masculine. In fact, the Netherlands scores much closer to the Femininity side of the spectrum (implying a focus on caring for others and the quality of life). Whereas Spain seems to balance the Masculine and Feminine characteristics, Italy is a highly Masculine society. Because in environments characterized by these masculine values, men are considered to have an advantage over women, men are expected to perform better (Ridgeway, 2001).[3] This suggests that the role of gender beliefs in the performance consequences of social-status ranking will be strong in Italy, absent in the Netherlands, and somewhere in between in Spain. Note that this order in 'masculinity' coincides exactly with the gender equality indices of the European Institute for Gender Equality (2017) that were discussed in the introduction.

Following the main reasoning of status characteristics theory outlined above, gender differences in performance will occur due to implicit status ranking. The differences in the Masculinity-Femininity cultural dimension suggest the following hypothesis:

Hypothesis 1. *When status ranking is implicit*, no gender differences in performance will be found in the Netherlands. Women will perform worse than men in Italy and Spain.

When status ranking is made explicit, Schram et al. (2019) report gender differences in performance for Spain. Although there is no theory to guide the effects of such explicitness, I conjecture that it strengthens the social evaluative threat. Given the country variation in the Masculinity-Femininity dimension, I therefore expect the same comparative outcomes as in hypothesis 1, but with stronger effects than in the implicit case.

Hypothesis 2. *When status ranking is explicit*, no gender differences in performance will be found in the Netherlands. Gender difference will be larger than with implicit ranking in Italy and Spain.

3. Experimental design and procedures

The same experiment was run in all three countries. The design is as follows. Each session consists of 13 participants, who are undergraduate university students. Upon arrival, they are randomly divided into six of type A, six of type B, and one of type C. The former two enter the laboratory, while the type C participant is brought to a separate room. The A-type participants are subjected to explicit status ranking, the B-type participants are not. Importantly, both A and B types can infer (implicit) status ranking from prior cultural beliefs about gender-specific performance, but only A-types participants are explicitly told they will be compared to others by a peer. Aside from culture (variation of the country in which the experiment takes place), this is the main treatment variation in the experiment. Both culture and the occurrence of explicit status ranking are varied between subjects.

The core of the experiment is that all A and B participants individually undertake a real-effort task, which is described in detail below. Their monetary earnings are based on their own private performance and are unaffected by any subsequent explicit social ranking (for the A types). For the B-type participants, individually undertaking a real-effort task is all that they do. Finally, C-type participants remain inactive throughout the experiment (see below).

The explicit status ranking of the A-type participants is organized as follows. Seated in a separate room is the randomly chosen C-type participant, that is, a peer. This C type knows nothing about the task that the A-type participants do and is only told that a higher score corresponds to a better performance.[4] Before they do the task, each A type is individually taken by a male or female experimenter (which was varied randomly) to the C-type participant. S/he enters the room (without the experimenter) and reads aloud a statement prepared by the experimenter. This states that s/he will return after the task to report her or his score on the task and how this ranks amongst the six A-type participants.[5] This first encounter between the A and the C type serves to create an anticipation of the explicit social ranking that will take place later. After completion of the cognitive task, the A types are once again taken to the C-type participant. Now they read aloud a text truthfully stating their score and their rank amongst the six A types.[6] Note that this makes the ranking explicitly *social*, because it is observed by a peer, the C-type participant.

This design allows us to compare the performance on the task of participants who anticipate explicit and implicit status ranking (type A) to those who anticipate only implicit status ranking (type B). The same design was applied in Barcelona (Spain), Amsterdam (the Netherlands), and Bologna (Italy).[7] All three experiments were conducted in a major city with a large student population (between 2000 in the Netherlands and 5000 in Italy), in experimental economics laboratories. This makes the subject pools used in the experiments similar on many dimensions outside of the cultural beliefs I am interested in.

The real effort task used was taken from Weber and Schram (2017). For 15 min, participants are presented with a sequence of pairs of 10 × 10 matrices filled with two-digit numbers. See Fig. 2 for an example of the computer monitor.

A participant must find the highest number in the left matrix and the highest number in the right matrix and calculate the sum of these two numbers. A correct answer yields one euro. This piece-rate remuneration is applied in all countries. It ensures that there is no rivalry

[3] In fact, Hofstede et al. (2010: 140) report that the names 'masculine' and 'feminine' reflect the degree to which gender roles overlap. For instance, in a masculine society "men are supposed to be assertive, tough, and focused on material success, whereas women are supposed to be modest, tender, and concerned with the quality of life". In a feminine society, both men and women are supposed to be the latter: caring and focused more on life quality.

[4] All participants are told that a good performance on the task has been shown to correlate positively with success in professional life. This was initially done in the Spanish sessions to create a belief about the importance of the social ranking. Subsequent analysis showed that exclusion of this text does not affect any of the results (as reported in Schram et al., 2019). Nevertheless, the text was also included in the Dutch and Italian sessions, to maintain consistency.

[5] See Appendix A for all instructions and read-out forms.

[6] One of the experimenters (who was isolated in a cubicle and did not know the participants, only their randomly determined desk ID) calculated their score and their ranking, and wrote it down in a piece of paper which was inserted in an envelope to ensure anonymity. This envelope was handed to the participant by a different experimenter. The participant was asked to read out aloud to the peer (C-type participant) the text written down by the first experimenter. The text was as follows: "Please read the following text to the participant waiting for you in that room: "My score on the task I did was … With this score, I was ranked … amongst the six participants." More details can be found in Appendix A.

[7] I thank the research groups BESlab at the University Pompeu Fabra, CREED at the University of Amsterdam, and BLESS at the University of Bologna for the use of their experimental economics laboratories. Note that use of these laboratories requires following their protocol. Deception is not allowed in these laboratories and debriefing is not part of the laboratory protocol. The design and procedures of the experiment were approved by the ethics committees in each of these laboratories.

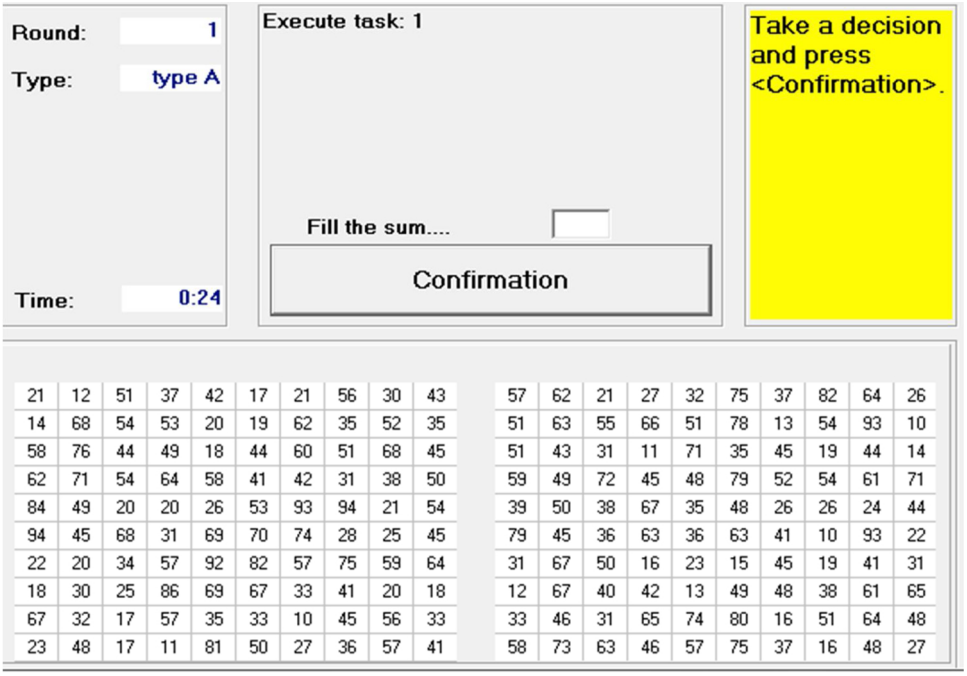

Fig. 2. Screenshot of real-effort task.
Notes. The instructions inform participants that the numbers in the cells were 'randomly generated'; a random mechanism was used that reduces the probability of high numbers compared to a draw with equal probabilities. This avoids a high probability of very high sums.

for resources (which would occur, for example, if only the best performer(s) were to be rewarded). Irrespective of whether or not the entered sum is correct, two new matrices appear.

The sessions were run in various waves in 2014–2016 (Barcelona), in 2015 (Amsterdam) and in 2017 (Bologna). Instructions were in Spanish (Barcelona), Dutch (Amsterdam) and Italian (Bologna). All participants were recruited on a voluntary basis from the respective subject pool of the three laboratories (cf. footnote 7). Participants were paid in euro in all cases. The experiment lasts approximately one hour and average earnings were €16.47, €22.44, and €15.91 in Spain, the Netherlands, and Italy, respectively. This includes a show-up fee of €7 in the Netherlands and €5 in Spain and Italy (if more volunteers show up than needed for a session, participants are randomly selected and the remainder is sent off with only the show-up fee).[8] I have 106 observations in Spain,[9] 71 in the Netherlands,[10] and 132 in Italy.[11]

4. Results

Fig. 3 shows the mean number of correct summations (which I interpret as the 'performance') per treatment, country, and gender. The figure also includes the numbers of observations per treatment cell and the results of permutation t-tests for gender differences (see Appendix C of Schram et al., 2019 for a detailed description of this test).

A first thing to notice in this figure is that without explicit status ranking, there are no significant gender differences in any country. This is in line with the prediction for the Netherlands in Hypothesis 1, but the lack of an effect for Spain and Italy provides no support for the status characteristics theory. When status ranking is implicit, namely when B-type participants do their task under ambiguous knowledge of their competence versus others but may infer relative competence from cultural gender beliefs, there are no gender differences between women and men's performances in any of the three countries. This finding also suggests that there is neither inherent nor 'status-inferred' gender difference in performance in this task. Further evidence is provided by a linear regression of the number of correct solutions on dummy variables indicating that the task was done in Italy, the task was done in the Netherlands (leaving Spain as the reference category) and a dummy indicating a female respondent. This regression is based on 179 observations. It shows that the Dutch are better at this task than the Spanish (the coefficient has a p-value below 0.01 and the effect size is 2.16 correct summations). The Italians are slightly worse than the

[8] Note that the marginal incentives are most relevant for this study and they were the same in each country (€1 per correct summation). Following the protocols of the laboratories involved, the show-up fee was higher in the Netherlands (€7) than in Spain and Italy (€5 euro). Upon advice from local researchers, it was decided to pay C-type participants less in Italy (€15 euro) than in the other two countries (€20). Note that this cannot affect the behavior of those I am interested in (A- and B-type participants) because they do not know the remuneration of C-types. Recall that the C-type participants were the passive players, who are not included in the analysis.

[9] In Barcelona, there were two participants with more than 150 entered solutions in the 15-minute time span. Without considering the matrices, they simply repeatedly entered the same number at a very high pace. Both had a very high number of correct summations. I drop them from further analysis. Because they were both males of type A, including them would further increase the gender difference documented below.

[10] In one of the sessions in Amsterdam, one of the participants of type B decided to leave before the end of the session.

[11] This variation in the numbers of observations across countries is driven by the participation rate in each country. Especially in the Netherlands, at the CREED laboratory in Amsterdam it has become an increasingly challenging task to recruit experimental participants. Furthermore, in each of the three cities, additional treatments were run to those reported here. This limits the number of participants who could be recruited for the treatments used in the present study. The data collected from the additional treatments are used in other research studies. Because these treatments were run in a between-subject design, using the data from (only) the sessions that were common in all three countries allows for the cross-cultural comparison I report here without bias from the other treatments. These additional treatments, used in other projects, are also the reason for the differences in the number of observations for the implicit versus the explicit treatments. More details about the other treatments in each country are available in Schram et al. (2019) for Spain and upon request for the Netherlands and Italy.

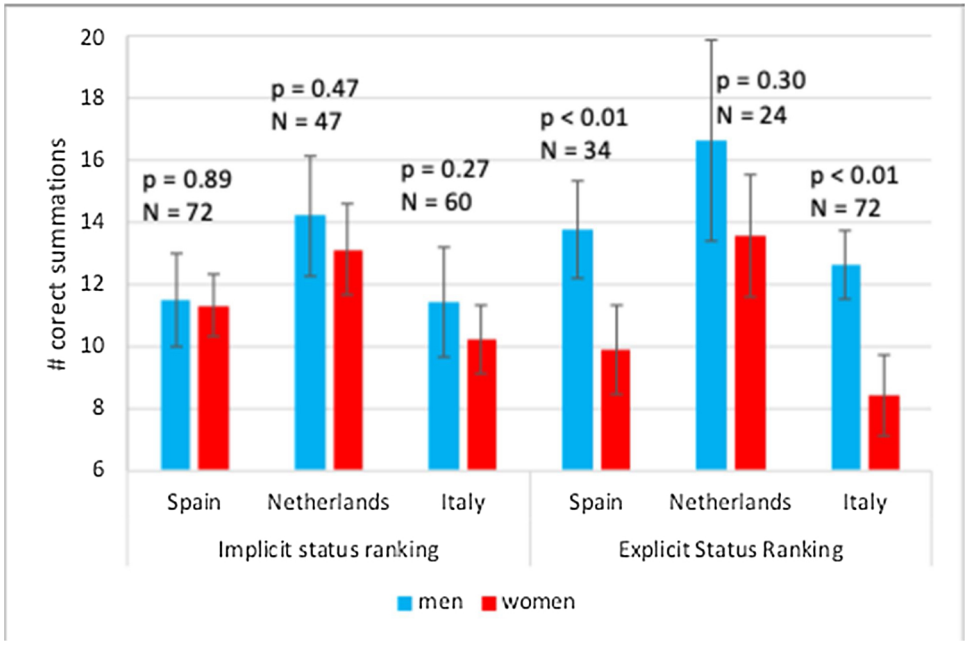

Fig. 3. Performance.
Notes. Bars indicate the mean number of correct summations per treatment-country combination; p-values report the result of a permutation *t*-test (Fischer test) for gender differences in the treatment cell concerned. Error bars denote 95 % confidence intervals.

Spanish (with an estimated 0.73 fewer correct summations), but the difference is statistically insignificant (p = 0.31). The gender dummy has a small effect (women have an estimated 0.80 fewer correct summations) and is statistically insignificant (p = 0.21).[12]

Turning next to the explicit status-ranking treatment, there are strong and significant gender differences in Italy and Spain, as hypothesized. In the Netherlands, the observed difference is small and statistically insignificant (also hypothesized).[13] Hypothesis 2 also predicts that in Italy and Spain, the gender difference is larger with explicit than with implicit ranking only. Indeed, I do observe larger and significant differences with explicit ranking in Spain and Italy, where no differences were found when only implicit ranking was at play. Hence, this part of Hypothesis 2 is also supported.

To further investigate the differential effects of explicit status ranking on men and women in each country, Table 1 reports the results of a linear regression of the number of correct summations on the interaction between gender and explicit status ranking. The regressions include also a dummy indicating that the participant has a major in business or economics. This serves as a control for possible differences in fields of study. The regression results for the interaction term confirm that when status ranking is explicit, women perform worse than men in Italy and Spain, whereas no significant gender differences are found in the Netherlands. Also, the effect sizes are much larger for Italy and Spain than for the Netherlands. On the other hand, the low values and insignificance of the coefficients for the female dummy confirm that there are no gender differences when there is only implicit ranking.[14]

Table 1
Gender differences in performance across countries.

	Spain		Netherlands		Italy	
constant	11.50	0.81**	13.73	1.25**	11.35	0.77**
Female	–0.17	0.94	–0.99	1.60	–1.10	1.02
Explicit Status ranking	2.28	1.16	2.54	1.60	1.21	0.99
Female x Explicit Status ranking	–3.73	1.54*	–1.94	2.83	–3.03	1.36*
Economics	0.00	0.81	0.79	1.27	0.67	1.37
N	106		71		132	

Notes. In each country, cells in the first column report the coefficients of linear regressions of the number of correct summations on the independent variables displayed in the rows; cells in the second column report the standard error terms. */** depicts statistical significance at the 5 %/1 %-level.

[12] As rightly pointed out by an anonymous reviewer, it is worth noting that there are some interesting gender differences in performance across the three countries. In particular, whereas the performance of Dutch women slightly increased when moving from implicit to explicit status ranking, the performance of Spanish and Italian women decreased. In order to certify whether such differences are large and significant, one needs to run a diff-in-diff analysis of gender differences across countries. This is an interesting approach, but it goes beyond the scope of this study due to the excessive numbers of observations that this requires.

[13] Using the Spanish results to conduct a power test to establish the numbers of observations needed in the Netherlands and Italy, shows that 80% power requires N = 24. This is the number of observations I have for the Netherlands, while for Italy I have 72 observations. Note that this power test is based on a t-test, whereas I use permutation t-tests. These are exact tests that do not make assumptions about the underlying distributions and as such have high power for as few as 8 observations per treatment cell (cf. Moir, 1998; Schram et al., 2019). For example, I have only 34 observations for Spain and observe a p-value of less than 0.01. Finally, I have 24 additional observations in the Netherlands where the explicit ranking is not social (participants each report their score –but not their ranking– to a different type C participant, but they do know how their rank amongst the A types). When pooling these observations with those represented in Fig. 3, the permutation t-test still shows no significant gender difference in the Netherlands.

[14] As rightly pointed out by an anonymous reviewer, the gender of the C-type participant may affect these results. The information on the gender of the C-type is only available for the Netherlands and Italy. I ran a regression of performance on the gender of A-type participant, the gender of C-type participant and their interaction. The effect of all these variables is small and insignificant in the Netherlands. In Italy, the effect of A-type's gender is large and strongly significant (confirming the main finding of the study, that compared to males, females underperform when there is explicit status ranking), whereas the effect of both, C-type's gender and its interaction with the A-type's gender, is small and insignificant.

5. Conclusion

This article examines the (differential) effects of status ranking on men's and women's performance. Social recognition of one's standing in a hierarchy, i.e., status ranking, has important implications for self and social assessment, and can as a result lead to (un)intended social stratification. Research inspired by the status characteristics theory (Correll & Ridgeway, 2003) has contributed to the main finding that in an uncertain environment where one's ability is unknown, being a man will infer a higher status than being a woman. This will subsequently lead to gender differences in self-assessment and eventually in individual performance. Gender and its categories –male/female– will serve as a status characteristic to infer performance expectations as long as there is a widely shared cultural belief that men are more competent at most things compared to women. However, status ranking is not made explicit in this line of research; it is implicitly inferred from cultural beliefs on whether men or women are expected to be better in a particular setting.

The first contribution of this article is that it theoretically recognizes and empirically tests the gender effect of status ranking when it is made explicit and compares this to an environment where it is only implicit. A laboratory experimental design allows for a proper investigation of the degree to which the gender effects of status ranking hinges on the implicitness or explicitness of this ranking. By randomly allocating one group of participants to an 'explicit status ranking' treatment and another group to no such treatment, the design allows us to compare the performance on the task of participants who anticipate explicit and implicit status ranking to those who anticipate only implicit status ranking.

A second contribution is that it examines the extent to which implicit and explicit ranking is culture dependent by considering three cultures that vary with respect to gender roles. This article is the first to offer a cross-country comparison of gender differences in the performance consequences of (implicit and explicit) social-status ranking. It does so by conducting the same experimental design in three culturally different countries, i.e., Italy, the Netherlands and Spain. According to Hofstede et al.'s (2010) cultural classification and the gender equality index of the European Institute for Gender Equality (2017), Italy and Spain are culturally closer to each other, than either is to the Netherlands. The cultural differences are especially pronounced regarding gender beliefs that are related to the Masculinity-Femininity dimension of this classification. Italy appears as a highly Masculine society, Spain balances Masculine and Feminine characteristics, and the Netherlands is the least Masculine. This means that Italy has a strong inclination towards competition and achievement, the Netherlands a strong focus on caring for others and quality of life, and Spain takes a position in between. In more competitive environments the social evaluative threat is likely to be strong and moreover, men are considered to have an advantage over women. The countries are therefore hypothesized to differ in the consequences of (implicit and explicit) social-status ranking for men's and women's performance.

The experimental results provide evidence that both the type of status ranking and culture matter. When status ranking is implicit, no gender differences in performance are observed. This is the case both when gender beliefs would not infer that men are better than women (like in the Netherlands), and when widely shared beliefs of men expected to be better are present (like in Italy and to a lesser extent in Spain). Hence, status characteristics theory does not find empirical support.

When status ranking is made explicit, strong gender differences are observed in Italy and Spain. The fact that this was not found in the Netherlands confirms that culture matters. In a society like the Netherlands, characterized by more feminine characteristics of cooperation and taking care of each other, the social evaluative threat may be lower and men's competence will typically be considered comparable to women's. As a consequence, no gender differences in performance are found, neither in the implicit nor in the explicit status ranking condition. In more masculine and competitive environments like the Italian and the Spanish, on the other hand, women do perform significantly worse than men. Importantly, however, cultural beliefs about men's and women's capabilities are not sufficient to drive gender differences when these beliefs are used to implicitly infer status ranking among men and women. Only when the status ranking (the own versus the others' competence) is explicitly given, do I find a gender effect. In other words, it seems that more is needed for gender inequality to kick in than simply inferring differences from status characteristics.

While such a conclusion may appear positive at first sight, the fact that there is a gender effect when the comparison is explicit remains a matter of concern. This is especially the case given the finding that men and women seem to be inherently similar in their abilities, that is, without the explicit status ranking, their performance in the individual task applied is very similar. The cross-country results suggest that cultural beliefs about gender are activated when participants anticipate an explicit comparison with others based on their performance. Since western societies are predominantly hierarchical, where performance ranking is widespread especially in labor market settings, women in the labor market will continue to remain in a disadvantage compared to men. The good news that stems from the results in the Netherlands is that this may change if widely shared cultural beliefs about gender become less biased against women.

Of course, these results are just a starting point. Obviously, more research is needed to establish the gender effects of explicit and implicit status ranking across different cultures. Hopefully, such research can build on the contributions of the present study. For instance, more countries can be studied, more data can be collected in the countries studied here (to increase the sample size) and alternative cognitive tasks can be performed. The individual task participants were asked to do was a summation task that some may experience as a male-oriented task.[15] I have no indication this is indeed the case (as part of a different project we even asked participants in another experiment and found no evidence that men are considered to be better at this task than women), but it would be interesting to examine whether the findings carry over to cases where the individual task might be viewed as more gender neutral or more female-oriented. Another interesting extension would be to run the experimental design in a same-gender environment. It is likely that the dynamics and underlying mechanisms will change compared to a mixed-gender environment like the one used here.

Finally, one might wonder whether there is a limitation related to the subject pool, which consists of (mainly) undergraduate students in the three countries. As a next step, the experimental design presented here could be conducted in the field.[16] A priori, there is no reason why gender effects in a field environment like a labor market setting would be different from those observed with undergraduate students in the laboratory. If anything, the fact that I do observe differences in the latter group, which is typically a more culturally homogenous group of participants, means that one might expect even stronger differences with a more representative sample of the population in each country. Thus one could say that the findings reported here are a conservative measure of the effect of culture on gender differences in performance. Indeed, Ridgeway (2014) summarizes numerous examples of cases, from the college student population to working populations, where gender is empirically shown to be used as a status characteristic, which

[15] Note that the main difficulty of the task is not the adding up of two two-digit numbers, but the focus that is needed to find the highest number in each of the two matrices.

[16] One possibility is to run the same experimental design in a workplace where workers' productivity is measurable. Then one treatment could be introduced where the workers know before doing their usual task that they have to report to a third party (e.g. a stockholder or public representative) their productivity and/or their ranking compared to other teammates' productivity. In the other treatment, workers are not required to report anything.

as a consequence disadvantages women. In this study, I preferred to take advantage of the control offered by a traditional laboratory environment before investigating the robustness of the findings in the field. I believe that the results reported here provide a sound basis for future explorations.

References

Anderson, C., Willer, R., Kilduff, G. J., & Brown, C. (2012). The origins of deference: When do people prefer low status? *Journal of Personality and Social Psychology, 102*, 1077–1088.

Ball, S. B., Eckel, C. C., Grossman, P. J., & Zame, W. R. (2001). Status in markets. *The Quarterly Journal of Economics, 116*, 161–188.

Berger, J., Cohen, B. P., & Zelditch, M. (1972). Status characteristics and social interaction. *American Sociological Review, 37*(3), 241–255.

Berger, J. M., Fisek, H., Norman, R., & Zelditch, M. (1977). *Status characteristics and interaction.* New York: Elsevier.

Bourdieu, P. (1984). [1979] In R. Nice (Ed.). *Distinction. A social critique of the judgment of taste*. Cambridge, MA: Harvard University Press.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *The American Journal of Sociology, 106*, 1691–1730.

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

Correll, S. J., & Ridgeway, C. L. (2003). Expectation States theory. In J. Delamater (Ed.). *The handbook of social psychology* (pp. 29–51). New York: Kluwer Academic/Plenum Publishers.

De Botton, A. (2004). *Status anxiety.* London: Hamish Hamilton.

Dickerson, S., & Kemeny, M. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychology Bulletin, 130*, 355–391.

DiMaggio, P. (1994). Culture and economy. In N. Smelser, & R. Swedberg (Eds.). *Handbook of economic sociology* (pp. 27–57). Princeton and N. Y: Princeton University Press and Russell Sage.

European Institute for Gender Equality (2017). *Gender equality index 2017: Measuring gender equality in the European Union 2005-2015 – Report*ISBN: 978-92-9493-768-1https://doi.org/10.2839/251500.

Falk, A., & Heckman, J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*, 535–538.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In (4th edition). D. T. Gilbert, S. T. Fiske, & G. Lindsey (Vol. Eds.), *The handbook of social psychology: 2*, (pp. 357–411). Boston: McGraw-Hill.

Gilbert, P. (1997). The evolution of social attractiveness and its role in shame, humiliation, guilt and therapy. *The British Journal of Medical Psychology, 70*, 113–147.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics, 118*, 1049–1074.

Heffetz, O., & Frank, R. (2008). Preferences for status: Evidence and economic implications. In J. Benhabib, M. Jackson, & A. Bisin (Vol. Eds.), *Handbook of social economics: 1A*, (pp. 69–91). Amsterdam: Elsevier.

Hofstede Insights (2019). *Country comparison.* https://www.hofstede-insights.com/country-comparison.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind*Revised and Expanded(3rd edition). New York: McGraw-Hill.

Merton, R. K. (1968). *Social theory and social structure.* New York: Free Press.

Minguez, A. M. (2010). Family and gender roles in Spain from a comparative perspective. *European Societies, 12*(1), 85–111.

Moir, R. (1998). A Monte Carlo analysis of the fischer randomization technique: Reviving randomization for experimental economists. *Experimental Economics, 1*, 87–100.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics, 122*, 1067–1101.

Parsons, T. (1970). Equality and inequality in modem society, or social stratification revisited. In E. O. Laumann (Ed.). *Social stratification: Research and theory for the 1970s* (pp. 13–72). Indianapolis, Ind: Bobbs-Merrill.

Paskov, M., Gërxhani, K., & van de Werfhorst, H. (2017). Giving up on the Joneses? The relationship between income inequality and status-seeking. *European Sociological Review, 33*(1), 112–123.

Ridgeway, C. L. (1991). The social construction of status value: Gender and other nominal characteristics. *Social Forces, 70*, 367–386.

Ridgeway, C. L. (2001). Gender, status, and leadership. *The Journal of Social Issues, 57*, 611–655.

Ridgeway, C. L. (2014). Why status matters for inequality. *American Sociological Review, 79*(1), 1–16.

Schram, A., Brandts, J., & Gërxhani, K. (2019). Social-status ranking: A hidden channel to gender inequality under competition. *Experimental Economics, 22*(2), 396–418.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *The American Psychologist, 52*, 613–629.

Stigler, G. (1987). Competition. In J. Eatwell, M. Milgate, & P. Newman (Eds.). *The New Palgrave: A dictionary of economics* (pp. 531). London: The Macmillan Press.

Veblen, T. (1931). *The theory of the leisure class.* New York: Random House.

de Waal, F. (1989). *Chimpanzee politics: Power and sex among apes.* Baltimore: Johns Hopkins University Press.

Weber, M. (1968). *Economy and society.* [1922]Los Angeles, CA: University of California Press.

Weber, M., & Schram, A. (2017). The non-equivalence of labor market taxes: A real-effort experiment. *The Economic Journal, 127*, 2187–2215.

Webster, M. A., Jr., & Foschi, M. (1988). Overview of Status generalization. In M. A. Webster, & M. Foschi (Eds.). *Status generalization: New theory and research* (pp. 1–20). Stanford, CA: Stanford University Press.

Wilkinson, R. G., & Picket, K. E. (2010). *The spirit level.* London: Penguin Books.