Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments

Katrin Auspurg,[a] Thomas Hinz,[b] and Carsten Sauer[c]

Abstract

Gender pay gaps likely persist in Western societies because both men and women consider somewhat lower earnings for female employees than for otherwise similar male employees to be fair. Two different theoretical approaches explain "legitimate" wage gaps: same-gender referent theory and reward expectations theory. The first approach states that women compare their lower earnings primarily with that of other underpaid women; the second approach argues that both men and women value gender as a status variable that yields lower expectations about how much each gender should be paid for otherwise equal work. This article is the first to analyze hypotheses contrasting the two theories using an experimental factorial survey design. In 2009, approximately 1,600 German residents rated more than 26,000 descriptions of fictitious employees. The labor market characteristics of each employee and the amount of information given about them were experimentally varied across all descriptions. The results primarily support reward expectations theory. Both men and women produced gender pay gaps in their fairness ratings (with the mean ratio of just female-to-male wages being .92). Respondents framed the just pay ratios by the gender inequalities they experienced in their own occupations, and some evidence of gender-specific evaluation standards emerged.

Despite women's rising participation in higher education in recent decades, gender inequalities in the labor market persist in the United States and other Western societies. Most prominently, women's hourly wages remain approximately 80 percent of men's wages (Blau and Kahn 2007; Ridgeway 2011). Even after extensively controlling for human capital and working hours, a remarkable gender pay gap remains, with the ratio of female-to-male earnings being .90 (for the United States, see Blau and Kahn 2007; for Germany, see Gartner and Hinz 2009). Moreover,

[a]LMU Munich
[b]University of Konstanz
[c]Radboud University Nijmegen

Corresponding Author:
Katrin Auspurg, Department of Sociology, LMU Munich, Konradstr. 6, DE-80801 Munich, Germany
E-mail: Katrin.Auspurg@lmu.de

policy efforts in recent years have not materially decreased gender inequalities (Akchurin and Lee 2013).

In this article, we argue that an important reason why notable gender pay inequalities persist is that both men and women perceive women's lower wages as fair, at least to a certain degree. There are several reasons to assume that wage distributions that observers consider fair will persist in labor markets. Following the *fair wage hypothesis*, workers proportionally reduce their efforts as their wages fall short of their perceptions of fair wages (Akerlof and Yellen 1990). Indeed, there is evidence that perceptions of unfairness cause low job satisfaction, low organization commitment, high turnover rates, absenteeism, and shirking (see, e.g., Colquitt et al. 2001). In the employer's perspective, perceptions of unfairness increase a firm's transaction costs, including searching for, bargaining with, and monitoring employees (Fehr, Goette, and Zehnder 2009; Husted and Folger 2004).

But why would women consider lower pay for women to be fair? To address this puzzle, we contrast two prominent explanations for just gender pay gaps. First, women might consider lower earnings to be fair because of *gender-specific referents*. Both the general tendency to compare oneself with others who are similar (Festinger 1954) and homophily in social networks (McPherson, Smith-Lovin, and Cook 2001) suggest that women compare themselves primarily with other women or with employees in female-typed occupations that have below-average wages. As a consequence, women might be partially unaware of the pay gaps between themselves and their male colleagues (Major 1989, 1994; Major and Forcey 1985). Second, women's inputs into the labor market might be perceived as being of less value than men's. This hypothesis of *gender status beliefs* is advocated by *reward expectations theory* (Berger et al. 1985; Berger, Wagner, and Webster 2014). Gender, as a diffuse status characteristic, is assumed to create cultural beliefs in men's higher competence and status worthiness, thus entitling them to higher rewards.

To date, little research contrasts these different approaches within the same research design. Such research, however, would be particularly promising for advancing knowledge on gender inequality and designing possible interventions. For instance, if just gender pay gaps are the result of gender-specific comparisons to others, attempts to reduce gender segregation in the labor market or to provide more information on actual wages would be helpful in reducing gender inequalities (Desmarais and Curtis 2001; Major 1989). From the perspective of reward expectations theory, however, increasing exposure to pay inequalities might instead promote the internalization of gender status beliefs. From this perspective, only additional measures (e.g., female role models earning high pay) would help destabilize gender inequalities (Ridgeway 2011).

We implemented two important innovations over prior research. First, using an experimental factorial survey approach, we varied the amount of information that described hypothetical employees. This ensured that gender differences in fairness evaluations of hypothetical employees' earnings were not caused simply by gender functioning as a proxy for lack of information on performance-related criteria, such as work experience (for such assumptions on statistical discrimination, see, e.g., Arrow 1998; Phelps 1972). Second, we collected information about respondents' occupations, thus enabling us to observe the extent to which fairness evaluations were framed by respondents' own social contexts. In 2009, approximately 1,600 residents of Germany participated in our experiments, resulting in more than 26,000 evaluations of fair earnings. Germany seemed well-suited for this research, because prior work has speculated whether subtle processes of discrimination could be the cause of the substantial gender pay gaps in Germany (see, e.g., Gangl and Ziefle 2009). This article contributes to the still underresearched question of the extent to which comparison processes and (status) beliefs play out not only in the lab but also in the real-world macro-context of societies.

THEORETICAL BACKGROUND AND EMPIRICAL EVIDENCE

Justice Evaluations

When analyzing how individuals make justice evaluations, researchers typically assume combinations of equality-, need-, and equity-based distribution rules (Deutsch 1985). The equality and need principles represent ideal standards, and the principle of equity is an existential standard in which the relation between rewards and inputs for rewardee A must equal the relation observed for a referent person or referent group B (Adams 1965; Shepelak and Alwin 1986; Walster, Berscheid, and Walster 1973). Just rewards are based on the typical reward for other people with similar reward-relevant characteristics, which means ideas about just rewards are linked to the existent distribution of rewards. For economic exchange situations, such as employment relationships, this existential standard is seen as being dominant, trumping ideal standards in their relevance (Shamon and Dülmer 2014).

Two theories explaining just gender pay gaps stand out. First, women might use other reference groups when making comparisons. Second, gender might be an input variable on its own, legitimizing higher earnings for people with higher status in this variable. This situation also includes the possibility of gender-colored evaluations of performance, as assumed in the extension of the *double standard hypothesis* (Foschi 1996, 2000; Foschi, Lai, and Sigerson 1994).

Gender-Specific Referents

Social comparison theories suggest that individuals prefer referents with similar characteristics, such as gender or occupation (Festinger 1954). One rationale is that these characteristics likely relate to performance and, as a result, provide ideal information for what one can expect (Gibson and Lawrence 2010). Another reason is opportunity structures: due to gender segregation in the labor market (Charles and Grusky 2004) and gender homophily in social networks (McPherson et al. 2001), male and female employees have different opportunities for comparison. Women typically work in sectors and professions characterized by lower wage levels (Charles and Grusky 2004). In combination with the use of same-gender referents, actual gender pay gaps translate into similar just gender pay gaps. In summary, one can first assume that women have generally lower pay expectations than do men. Second, gender differences in fair wages mirror actual pay gaps, and third, this equivalence exists, in particular, in areas where women have little contact with male co-workers (i.e., in highly segregated occupations).

Research on social comparison processes confirms that subjects compare themselves with similar others (Bylsma and Major 1994; Major 1994; Major and Testa 1989). In several experiments, women first showed lower pay expectations than men, but these differences disappeared once both genders were provided the same comparison information (Bylsma and Major 1992, 1994; Major, McFarlin, and Gagnon 1984; Major and Testa 1989). Unfortunately, the authors of these laboratory studies were not able to test whether these effects are generalizable beyond the gender-neutral tasks and student participants they used.

Convincing evidence of the same-gender referent hypothesis would also have to show that women hold lower reference standards not only for themselves but also for their male colleagues. Few studies test this notion. Major and Konar (1984) found that female management students estimated the typical pay of managers (independent of their gender) as lower than did their male classmates, and this difference was one of the main reasons why female students held lower pay expectations. However, as the authors stated, the participants in this study were asked to estimate typical wages in the fields they planned to enter, which means women likely anticipated wages in female-dominated areas such as personnel management.

Gender as a Status Value: Reward Expectations Theory

Reward expectations theory is rooted in status characteristics theories and the status value theory of distributive justice (Berger et al. 1977; Melamed 2012). These approaches attempt to explain how status structures emerge via social interactions and how nominal characteristics such as gender become connected to different (reward) expectations (for overviews, see Berger et al. 2014; Berger and Webster 2006; Kalkhoff and Thye 2006).

Reward expectations theory, in particular, focuses on explaining how expectations about fair rewards are activated within an actor's encompassing social framework (Berger and Webster 2006). A core assumption is that justice evaluations require stable frames of reference: "reward expectation inputs include not simply task related contributions within the local setting, but also culturally relevant norms regarding the social worth of different types of actors" (Fişek and Hysom 2008:771). These referential standards provide socially shared beliefs about how rewards are typically allocated within groups or a society as a whole (Berger et al. 1985; Fişek and Hysom 2008). Individuals who use these referential standards come to expect the same compensation for their services as that provided to people who show the same (status) characteristics (Berger et al. 1985; Shepelak and Alwin 1986). Additionally, when a socially valued reward is distributed unequally, actors infer respective performance differences based on these reward differences (Berger et al. 1985).

Berger and colleagues (1985) distinguish three types of referential structures that might be activated in social comparison processes: (1) abilities—what actors can do in a situation; (2) performance—what actors have accomplished; and (3) categorical information on status characteristics—that is, who the actors are. Status characteristics can be either specific or diffuse: specific status characteristics carry widely shared expectations for competence in limited, well-defined ranges of ability, whereas diffuse status characteristics carry very general expectations of higher competence (Berger et al. 1985; Correll and Ridgeway 2003). Gender is considered a diffuse status characteristic, that is, men are commonly expected to be more competent in most tasks, or at least tasks that "count most" (e.g., tasks requiring instrumental rationality, management tasks; Correll and Ridgeway 2003; Thébaud 2015).

To summarize, status beliefs are socially shared beliefs that consensually value one category of a social attribute as more worthy and competent than another, which means people who are disadvantaged by status beliefs accept such beliefs (Correll and Ridgeway 2003). This constitutes the most important difference to the same-gender referent hypothesis: one should expect not an effect of the *observer's* gender (i.e., female observers consider lower base wages to be fair), but rather an effect of the *rewardee's* gender: *both* male and female observers are assumed to believe in the greater competence of male employees and thus to assign higher just earnings to male employees.

As a highly visible status marker, gender should be salient in most social situations (Melamed 2012; Ridgeway 2011). However, variations in effect sizes may exist. On the one hand, gender may be particularly salient when actors of different genders frequently interact with one another, that is, in gender-mixed occupations (Rashotte and Webster 2005; Thébaud 2015). On the other hand, gender status beliefs that give men more credit might be more salient in male-stereotyped occupations (Berger and Fişek 2006; Ridgeway 2011). In such occupations, in addition to diffuse beliefs about male competence, individuals also have presumptions about gender-specific skills (Ridgeway and Correll 2004).

Gender status beliefs should also be particularly dominant in social contexts in which gender inequalities already exist: a core assumption of reward expectations theory is that existing inequalities are reproduced. One can thus assume that fair wage gaps are shaped by actual pay gaps in observers' and

rewardees' occupations. Both seem to be meaningful reference points in third-party evaluations.[1]

Finally, another assumption is built on the combination of reward expectations theory with theories of double standards: status characteristics are assumed to prime not only individuals' expectations of competence and performance, but also the standards used to interpret information about performance as suggestive of true abilities (Foschi 1996, 2000). Status inconsistencies, such as low-status group members showing high performance, can be reconciled with existing status beliefs by using a harsher standard. As a consequence, status beliefs should cause biased assessments of performance. People with a disadvantaged status (women) are held to a stricter standard unless their performance is judged as being "good" (Foschi 1996, 2000; Thébaud 2015).

Research on expectation states theories builds on a long tradition of using standardized experimental settings.[2] Dozens of laboratory studies have found support for most of the proposed mechanisms (for overviews, see Berger et al. 2014; Berger and Webster 2006; Kalkhoff and Thye 2006). Although one must conduct research in non-laboratory settings to determine whether status beliefs are also primed by real-life experiences or can be generalized to real (labor market) settings, few studies do so.

One of the few exceptions is the seminal field experiment conducted by Correll, Benard, and Paik (2007). The authors reported a motherhood penalty, in the sense that mothers are less likely to be invited to a job interview than are equally qualified men or non-mothers. This finding alone might also represent evidence of statistical discrimination (more on this theory will be presented later). Only in combination with a factorial survey experiment were the authors able to provide more direct evidence of status beliefs. University students reported that mothers should demonstrate higher test scores in management abilities before being considered as hirable as non-mothers (Correll et al. 2007).

This additional evidence for double standards allowed the authors to conclude that status beliefs had influenced respondents' answers. However, this evidence was again based solely on university students.

This restriction is also true for Thébaud (2015), a study in which university students rated vignettes showing fictitious descriptions of entrepreneurs. Again, participants held lower performance expectations for women and rated their entrepreneurial abilities and business plans more harshly. These gender differences were more pronounced in settings in which entrepreneurship was male-typed (e.g., in high-tech instead of a gender-neutral industry; and more so in the United Kingdom than in the United States, with men's over-representation in entrepreneurship actually being stronger in the United Kingdom).

In addition, some factorial survey studies have asked respondents to evaluate the fairness of earnings of hypothetical employees, which is very close to the empirical design used in the current study. Studies using general population surveys consistently find evidence of a just gender pay gap favoring men (for Germany, see Sauer et al. 2014; for Switzerland, see Jann 2005; for Ukraine, see Gatskova 2013; for the United States, see Jasso and Webster 1997). Jasso and Webster (1997) fit new theoretical models to the factorial survey module originally designed and administered by Jasso and Rossi (1977) to a blocked-quota probability sampling of 200 white adults in Baltimore in 1974. They also examined underlying mechanisms and found evidence of a fair base-wage gap (men were generally assigned higher earnings) in combination with double standards for education (men in the vignettes were assigned higher fair returns on education). However, as the authors stated, the study was designed primarily to illustrate the potential application of factorial surveys rather than to study causal mechanisms.

Interestingly, factorial surveys with university students—or with an overrepresentation of highly educated, young respondents—did not find just gender pay gaps or even slight evidence of just gender pay gaps favoring

women (for Europe, see Auspurg, Hinz, and Liebig 2009; Shamon and Dülmer 2014; for the United States, see Jasso and Webster 1999). To date, it is unclear whether these discrepancies represent age or cohort effects. Younger cohorts might share more egalitarian pay norms. Another explanation, which has not yet been considered, is that university students, even those who have experienced gender roles or specific status beliefs of women (men) being more competent in typical female (male) tasks, might still lack the exposure to (pay) inequalities in work settings that fuel diffuse status beliefs of male workers' *generally* higher competence.

A Note on Differential Norms and Statistical Discrimination

In justice theory, not all reward-relevant characteristics have to possess status value; for instance, observers might also assume higher rewards go to people with higher financial needs (Jasso and Webster 1999). A classic assumption was that women might rely more heavily than men on ideal standards (need and equality) instead of the equity principle (Davison 2014). As a consequence, women would demand relatively lower remunerations for status characteristics such as labor market achievements (e.g., educational investments and abilities). Yet this *differential norm hypothesis* found little empirical support (Mueller and Kim 2008). We thus do not discuss this hypothesis in more detail, although we will test it.

Theories focusing on stereotypes regarding probabilistic distributions of group characteristics are not directly concerned with justice evaluations, but they provide a further mechanism for individuals to discriminate against members of their own group (for theories on statistical discrimination, see Arrow 1973; Fang and Moro 2011; Phelps 1972; Schauer 2003). Their core argument is that group membership serves as cheap information on characteristics that are difficult to observe. Gender is considered a proxy for labor market productivity, at least insofar as it is likely connected to work effort in general (e.g., because women's likely responsibility for housework causes them to expend less effort on work outside the home; see Becker 1985). Furthermore, work interruptions for family reasons result in higher turnover costs, particularly in occupations with high training costs (Bielby and Baron 1986; England 1992). In addition, on average, mothers take more days off for family needs, such as children's illnesses (Nielsen, Simonsen, and Verner 2004).

Therefore, lower earnings for women can represent adequate compensation for lower performance, indicating lower productivity on average (statistically). However, given full information on a worker's ability and performance, one would no longer expect differences in evaluations of male and female workers; with increasing information on performance, discrimination should decline. This notion is in direct contrast to the argument that predicts gender status beliefs to be immune (or at least very resistant) to information on other input variables, such as performance or ability. Following this theory, gender stereotypes are largely buffered "from the immediate impact of disconfirming gender experiences" (Ridgeway 2011:186).

Several scholars report evidence of statistical discrimination, or the *proxy hypothesis*, based on factorial survey experiments (for the labor market, see Jann 2005; for the housing market, see, e.g., Krysan et al. 2009).[3] But none of these studies test the core mechanism (demographic categories or neighborhood characteristics used as proxies); hence, much of what is interpreted as evidence of statistical discrimination might actually be evidence of status beliefs. Ruling out the mechanism of status beliefs requires variation in the amount of information.

Summary on the State of Research

All strands of the theories discussed above assume a gender bias that disadvantages women in such a way that conceptions of "what is" become the basis of "what ought to be" (Homans 1973; Shepelak and Alwin

1986). All seem plausible and have some empirical support, at least when tested separately, which might, however, have led to misinterpretations. For instance, some survey studies interpret the lower pay expectancies of groups with lower status value (women) as evidence of reward expectations theory (Correll and Ridgeway 2003; Moore 1991). This finding alone, however, would also be perfectly in line with the same-gender referent hypothesis. Similarly, the observation that women likely compare themselves with other women is cited as evidence of the same-referent hypothesis, but it would likely fit even better with reward expectations theory: if gender is a status characteristic, using same-gender referents is more informative than using other referents with other status values. Gender status beliefs could simply be the motivational factor driving the tendency to use same-gender referents.

More powerful contrasts among the different causal mechanisms would be of academic and practical value. From the perspective of the gender-specific comparison, desegregating male and female employees, or ensuring more transparency in pay, would help reduce gender inequalities in (fair) pay. From the perspective of reward expectations theory, these measures would not be sufficient and might even be harmful: following this line of reasoning, the only way to eliminate status beliefs would be to frequently present individuals with role models of high-status women being rewarded with high pay.

RESEARCH STRATEGY

In the following section, we demonstrate that the multiple standard framework proposed by Jasso (1980, 1996) and its formalization in the "justice evaluation function" can provide an appealing way to contrast the different mechanisms within one coherent design: the same-gender referent hypothesis and the hypotheses on gender as status value. In addition, we discuss predictions by gender-specific double standards, the differential norm hypothesis, and possible statistical discrimination.

Jasso's justice theory suggests how feelings of injustice are related to just and actual levels of rewards (earnings). The main proposition formalizes the justice evaluation J as a logarithmic function of the ratio of actual (A) and just rewards (C), multiplied by an expressiveness constant θ ("theta"), which leads to the following "justice evaluation function":

$$J = \theta \ln\left(A/C\right) = \theta \ln\left(A\right) - \theta \ln\left(C\right) \qquad (1)$$

In the case of perfect justice, A equals C, and the evaluation J takes the value 0; in the case of under-reward, $C > A$, and J takes negative values indicating "unfairly low"; over-reward is signaled by positive values of J. This formula allows one to determine the amount of fair wages C (or fair gender pay ratios) when individuals' justice evaluations J and the rewardee's actual earnings A are known.

In our study, we investigate respondents' justice evaluations of fictitious employees sketched in vignettes (more details in the Methods section). Jasso's justice evaluation function implies that these justice evaluations are a linear function of the (status) characteristics the rewardee or vignette person i shows, including, for instance, education (X_{educ}), number of children (X_{child}), and gender (X_{female}), as well as logarithmic actual earnings, $\ln(A)$. These different input criteria enter a linear regression function explaining the justice evaluation J_{ij} of observer (respondent) j (see Jasso and Webster 1999:372; for details, see the online supplement [http://asr.sagepub.com/supplemental]):

$$J_{ij} = \alpha_j + \beta_{educ} X_{educ_i} + \ldots$$
$$+ \beta_{female} X_{female_i} + \beta_{\ln(A)} \ln\left(A_i\right) + \varepsilon_{ij}$$

with $i = 1, \ldots, n_{rewardee}; j = 1, \ldots, n_{observer}$ (2)

The values of the regression constant α_j represent—given the covariates—the conditional base level of how observers j evaluate the wages of the rewardees; ε_{ij} represents the residual error of the regression model (see the Method section for more details). Positive (negative) coefficients β of the different input variables (vignette variables) indicate that a

rewardee is over- (under-)rewarded in comparison to reference rewardees not showing these attributes, and the size of the different regression coefficients provides information on the relative weight of the input variables. For instance, one can expect that having much labor market experience or high education entitles rewardees to higher rewards (pay), which is reflected by negative regression coefficients (Jasso and Rossi 1977; Jasso and Webster 1997).

The regression equation can easily be extended using characteristics of the observers that we denote with the letter Z (e.g., their gender) or cross-level interactions between observers' and rewardees' characteristics ($X \times Z$). We can use those interactions to test whether the impact of rewardees' characteristics is moderated by observers' characteristics (e.g., a significant effect of the interaction term $X_{female} \times Z_{female}$ would indicate that male and female observers differ in their gender status beliefs). Observers supporting earnings equality should not be influenced by rewardees' characteristics, but instead should react more strongly to low- or high-earning values $\ln(A_i)$, meaning they show a particularly strong effect size $\beta_{\ln(A)}$, which can be tested by cross-level interactions of observers' variables with $\ln(A_i)$. Similarly, we can use interactions between input variables to test whether observers evaluate them differently for male and female rewardees (for more details, see Jasso and Webster 1999 and the Methods section).

In summary, this research strategy seems to have achieved precisely what is important in advancing our knowledge: differential inputs, norms, and referents can be analyzed within one coherent framework (Mueller and Kim 2008). Table 1 shows how the assumptions made in the different theories can be re-specified in this framework.

In addition, this framework allows us to derive illustrative information on the size of gender pay ratios that respondents consider fair (Auspurg and Hinz 2015; Jasso 1990). This information is not necessary to test our hypotheses but is indicative of the practical relevance of results (effect sizes). We use gender multipliers that are based on regression estimates when regressing the fairness evaluations J_{ij} on rewardees' characteristics, including their gender and log-earnings $\ln(A_i)$. These multipliers represent the ratio of mean just pay for women compared to men, adjusted for the vignette variables controlled in the regression. For instance, a multiplier of .8, estimated by a regression including vignette persons' educational degrees and occupations, suggests that, following the observers' fairness ratings, the mean fair pay for women should be .80 times (80 percent) that of men with the same educational degrees and occupations. We obtained sample estimates of these just gender pay ratios (JGPRs) by taking the exponential function of the negative ratio of the gender coefficient to log-earning coefficient: $JGPR = \exp(\beta_{female}/-\beta_{\ln(A)})$. (For more details, including on standard errors, see the online supplement.)

METHODS AND DATA
Experimental Design

On one hand, to test the hypotheses outlined earlier, one needs an experiment that reveals the causal influence of gender on justice evaluations. On the other hand, we are interested in how the structural contexts, in which observers are embedded, shape the influence of gender. Factorial survey (FS) designs combine experimental research with the possibility of including broad population samples (for introductions, see Auspurg and Hinz 2015; Mutz 2011). Within FSs, respondents are asked to evaluate short descriptions of hypothetical objects or situations (vignettes). These vignettes consist of several attributes (dimensions) that are experimentally varied in their levels. Researchers can thus disentangle how individual dimensions influence respondents' evaluations, and with the inclusion of the experiment in a survey, one can easily realize respondent samples that vary in labor market experience and other social contexts.

This approach is often called a "multiple standard framework" (Jasso and Webster

Table 1. Overview of Theoretical Assumptions and Implications for Model Parameters

Theory	Hypotheses	Expectations for Regression Coefficients Belonging to the Following Input Variables[a]
Same-gender referent hypotheses	1.1 Female *observers* in general consider lower base wages for both male and female rewardees to be fair (i.e., they more likely rate vignette persons as overpaid)	$Z_{female} > 0$
	1.2 This is particularly true in the case of high gender segregation in *observers'* occupation	$Z_{female} \times Z_{occ\_f} > 0$ (with Z_{occ_f} indicating the observer is working in a female-dominated occupation)
Reward expectations hypotheses	2.1 Male and female observers consider lower earnings for female *rewardees* as fair	$X_{female} > 0$
	2.2 This is particularly true when the *rewardee* works in a male-dominated occupation	$X_{female} \times X_{occ\_m} > 0$ (with X_{occ_m} indicating a male-dominated occupation)
	2.3 Fair pay ratios are shaped by actual gender pay ratios in *rewardees'* and/or *observers'* occupations	$Z_{payratio} \times X_{female} < 0$ $X_{payratio} \times X_{female} < 0$ (with $Z_{payratio}$ and $X_{payratio}$ denoting the actual gender pay ratios in observers' resp. rewardees' occupations)
	2.4 *Extension in form of double standards hypothesis*: Indications of a performance above (below) average are more strongly accounted for male (female) rewardees	$X_{perform\_above\_average} \times X_{female} > 0$ $X_{perform\_below\_average} \times X_{female} > 0$ (with $X_{perform\_above\_average}$ resp. $X_{perform\_below\_average}$ denoting a high resp. low performance input of the rewardee)
Additional Tests		
Statistical discrimination	3 Adding information on abilities and performance reduces the impact of rewardees' gender	$X°_{female} < X_{female}$ (with $X°_{female}$ denoting the effect of rewardees' gender in a situation with more information on this rewardee compared to the situation of X_{female})
Differential norm hypothesis	4 Men more likely hold norms of equity; women more likely hold norms of need and equality (i.e., they support lower rewards for education and work experience but higher rewards for children, and they are more sensitive to the earnings values presented in the vignettes)	$Z_{female} \times X_{educ} > 0$ $Z_{female} \times X_{exp} > 0$ $Z_{female} \times X_{children} < 0$ $|Z_{female} \times \ln(A)| > 0$

Note: Z variables belong to observers; X variables denote characteristics of rewardees.
[a]For the sake of brevity, we skipped the regression coefficients throughout the notations in the table. We may explain this with one example: Assumption 4 of the differential norm hypothesis is that the regression coefficient belonging to the cross-level interaction of female observers with highly educated rewardees ($Z_{female} \times X_{educ}$) is positive, meaning that female observers rate higher-educated rewardees as being less underpaid compared to male observers.

1997, 1999; Mueller and Kim 2008): observers (respondents) judge rewardees, and researchers can study how characteristics of observers *and* rewardees affect expectations of (fair) rewards. In studying justice evaluations, such samples offer much firmer ground than do convenience samples (e.g., the selective student samples used in most laboratory experiments on same-gender referent and reward expectations theories). In addition, FS designs are one of the few methods that enable mostly unbiased discrimination research (Mutz 2011; for a more critical assessment, see Pager and Quillian 2005).

We used the FS method to analyze whether respondents use different evaluation standards for male and female employees and, if so, under what conditions. In our FS module, respondents were asked to evaluate the fairness of the monthly gross earnings of fictitious full-time employees who were described by dimensions that indicated primarily different labor market characteristics. Most importantly, male and female vignette persons were described (on average) using exactly the same labor market characteristics; the gender of the vignette person was not correlated with any of the other vignette dimensions.[4]

Respondents were randomly assigned to sets of 10, 20, or 30 randomly ordered vignettes (we varied the number of vignettes per respondent to check for possible fatigue effects in bigger vignette sets; see Sauer et al. 2011). We experimentally varied the amount of information provided about these fictitious employees by using a between-subjects design with three different splits consisting of 5, 8, or 12 dimensions on labor market-relevant information. These splits were randomly allocated to respondents and allowed us to test whether the vignette character's gender was used as proxy information (statistical discrimination) in low-information conditions, or whether respondents were immune to extensive information on labor market performance (which would support the assumption of status beliefs).

Overall, 1,604 respondents evaluated 26,207 vignettes; 528 respondents evaluated the 5-dimension vignettes (8,698 judgments), 531 respondents evaluated the 8-dimension vignettes (8,717 judgments), and 545 respondents evaluated the 12-dimension vignettes (8,792 judgments).[5] Whereas the vignettes with five dimensions provided information only about employees' age, gender, education, occupation, and gross earnings, in the other splits additional information about labor market experience, tenure, and further characteristics of employees and their work settings was provided (see Table 2). Statistical analyses (see below for more information) show that these experimental splits were balanced in terms of the core vignette dimensions and respondents' characteristics, including their labor market experience, age, and gender.

Figure 1 shows a sample vignette for the 8-dimension split. Respondents rated the fairness of earnings on an 11-point scale from −5 (unfairly low) to +5 (unfairly high). All the dimensions were selected according to labor market research and previous evidence from FSs on the fairness of earnings (Sauer et al. 2011). We used qualitative levels (e.g., "little" versus "much") for the dimensions *labor market experience* and *tenure*. We did so despite the methodological literature suggesting that respondents might interpret such "vague quantifiers" in idiosyncratic ways or according to standards that might differ across the different (gender) groups under evaluation (Neumark 2012; Schaeffer 1991). Quantitative levels would have led to many implausible combinations (e.g., a 30-year-old university graduate having 20 years of labor market experience). Simply dropping those combinations would have destroyed the unrelatedness (orthogonality) of the experimental vignette dimensions. Extensive checks for unobserved heterogeneity did not reveal any hints that these vignette levels were interpreted in a gender-specific way (e.g., using different anchors for "much" labor market experience when evaluating female instead of male vignette persons). Further limitations of our choice of vignette levels are discussed subsequent to the presentation of results.

Table 2. Vignette Dimensions and Levels

#	Dimensions	Levels
1	Age	30/40/50/60 years
2	Gender	Male/female
3	Vocational degree	Without degree/vocational training/university degree
4	Occupation	Unskilled worker/door(wo)man/engine driver/clerk/ hairdresser/social worker/software engineer/electrical engineer/manager/medical doctor
5	Gross earnings/month	Ten values ranging from 500 to 15,000 euros
6	Experience	Little/much
7	Job tenure	Entered recently/entered a long time ago
8	Number of children[a]	No child/1 child/2 children/3 children/4 children
9	Health status[b]	No health problems/ health problems for a long time
10	Job performance	Below average/average/above average
11	Economic situation of the organization	High profits/threatened by bankruptcy/solid
12	Organization size	Small/medium/large

[a]The category *no child* was oversampled to achieve a more realistic distribution of family size.
[b]The category *no health problems* was oversampled for the same reason.

We built the vignette sample of 240 different vignettes using a fractionalized experimental design (*D*-efficient design), which is state of the art for multifactorial survey experiments (Auspurg and Hinz 2015). *D*-efficient designs are built by a computer algorithm that searches for a sample characterized by a minimal intercorrelation of dimensions (and interaction terms) and a maximal variance of vignette levels. This ensured that vignette dimensions and interaction terms are mutually uncorrelated. In addition, the design features lead to minimal standard errors in regression estimations and, therefore, maximum statistical power in revealing their impact on evaluations.[6] The sampling excluded illogical cases (e.g., implausible combinations of education levels, earnings, and occupations; for details, see Sauer et al. 2011). We also used *D*-efficiency as target criterion to allocate the vignettes to the different questionnaire versions (*decks*) consisting of 10, 20, or 30 vignettes. The sampling technique ensured there was no correlation between vignette persons' characteristics and especially that there would be no correlation between their gender and gross earnings (in the realized sample, this correlation was $r = .01$). In other words, there exists no gender pay ratio in the vignette sample.

This design has important advantages. First, the FS method enabled us to create what is rarely observable within real labor market data: male and female employees who are characterized by the same (labor market) attributes. This standardization makes it possible to observe whether gender is really a status value in itself, or whether it is only a proxy for missing information on job performance. Second, the indirect question format of FS methods enables the disclosure of very subtle discriminatory attitudes that are unconscious to respondents or prone to social desirability bias (Mutz 2011). This is essential when researching status beliefs that arise mostly out of actors' awareness (Ridgeway 2011). Third, the random allocation of vignettes ensured that all respondents (all men and women in the sample, individuals working in typical male or female occupations) were presented, on average, the exact same scenarios. Only this separation of vignette cases from respondents' labor market experience enables us to study the causal

A 50-year-old woman with no vocational training has two children. She works as a clerk and has a lot of job experience. She has worked for the organization for a long time.
Her monthly gross earnings total 1,200 euro (before taxes and extra charges).

Are the monthly gross earnings of this person fair, or are they, from your point of view, unfairly high or low?

Unfairly low					Fair					Unfairly high
−5	−4	−3	−2	−1	0	+1	+2	+3	+4	+5
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 1. Sample Vignette with Eight Dimensions

effect of gender priming in occupations on evaluative beliefs. Most of the vignettes were evaluated (97.3 percent), which suggests that respondents felt able to rate the fairness of the fictitious employees' earnings (see also the robustness checks at the end of the Results section).[7]

Respondent Sample

This study's data were collected in a general population survey (*n* = 1,604 respondents) in Germany in 2009. Approximately half of our 1,604 respondents who made valid vignette judgments were female, with a mean age of 49 years; 37 percent worked full-time, 16 percent were part-time or were marginally employed, and 47 percent did not participate in the labor market. Approximately 21 percent of the respondents lived in East Germany, which matches the shares in the population. Germany is a conservative welfare state with traditional patterns of work sharing between the sexes, as manifested in the much higher proportions of women than men who either work part-time or take parental leave.[8] There are additional, pronounced gender inequalities in the German labor force, including strong earnings differentials. Overall, the ratio of female-to-male hourly wages was .77 (Eurostat 2015). Controlling in regression analyses for observable covariates, such as labor market experience, industries, and occupations within firms ("job cells"), the "adjusted" gender pay ratio was .88 (Gartner and Hinz 2009). In addition, the German labor market is highly gender segregated

(Charles and Grusky 2004). Compared to many other European countries, including traditional welfare states such as Italy, and in particular compared to the United States, Germany has extensive entitlements for parental leave and generous social policy provisions for parents (see, e.g., Aisenbrey, Evertsson, and Grunow 2009) that likely enforce traditional, gender-specific work arrangements and thus provide strong incentives for statistical discrimination (Gangl and Ziefle 2009) or status beliefs. Accordingly, Germany is an ideal context to study how recipients' gender affects justice evaluations. Moreover, there are substantial differences regarding payment between East and West Germany. For instance, in 2009, the unadjusted ratio of female-to-male hourly wages in West Germany (estimated using data from the German Socio-Economic Panel) was .75, whereas in East Germany it was .93 (Smolny and Kirbach 2011). This variance provides further opportunities to test the extent to which social contexts shape justice evaluations.

Respondents were sampled to represent the total (residential) adult (over 18 years of age) population of Germany (for detailed information on all technical features, see Sauer et al. 2011). We used a mixed-mode approach (computer-assisted personal interview [CAPI], computer-assisted self-interview [CASI], and paper-and-pencil self-interview [PAPI]) to assess the sample. Response rates were between 48 percent (random walk for CAPI) and 60 percent (respondents who answered the CASI/PAPI after being successfully recruited via random-digit dialing). Note that similar to

other experiments, FSs do not necessarily require representative respondent samples to ensure internally valid results (Mutz 2011). The analysis sample contained 777 respondents surveyed with the CAPI mode, 435 respondents who responded via the CASI mode, and 392 respondents who used the PAPI mode. All survey participants were required to respond to the FS module and answer detailed questions about their own sociodemographics and labor market experience.

In addition to this survey, data to assess the proportions of female employees and actual gender pay ratios within different occupations were needed. The German Socio-Economic Panel Study (SOEP) served as a suitable source; it covered 11,370 employed individuals in 2009 (for general information on this panel, see Wagner, Frick, and Schupp 2007).[9] We calculated the female-to-male gender pay ratios based on employees' hourly wages (excluding hourly wages lower than two euros) by occupation using three-digit ISCO codes, each based on the average of 132 observations. The arithmetic mean of the gender pay ratio for respondents' occupations was .84 with a median of .82. In 91.5 (8.5) percent of respondents' occupations, the gender pay ratio was lower (higher) than 1, indicating that women's average pay was lower (higher) than men's; no occupation had perfect gender equality (pay ratio of 1). About 60 (16) percent of female (male) respondents were working in female-dominated occupations (at least 70 percent females), 35 (47) percent in integrated occupations, and 5 (37) percent in male-dominated occupations. The arithmetic mean of the actual gender pay ratios for the 10 vignette occupations was .87 with a median of .82 (for more information on the distributions of respondents' variables and gender pay ratios, see Part D of the online supplement). We matched data on actual gender pay ratios using the occupational codes of vignette persons (VPs) and respondents (respondents' actual occupation; if respondents were not employed at the time of the survey, we used codes for their last part- or full-time occupation).

We employed multiple regression techniques to estimate how vignette dimensions influenced respondents' judgments. Applying the common formulas for justice evaluations (Jasso 1996), we used the logarithmic specification of the wages provided in the vignettes.[10] To control for the impact of occupations, we included occupational prestige, measured by the Standard International Occupational Scale (SIOPS; Ganzeboom and Treiman 1996). The data structure is hierarchical: each respondent evaluated several vignettes, so ordinary least square (OLS) regression models show biased standard errors (Hox, Kreft, and Hermkens 1991). To account for this data structure, we used multilevel regressions (random intercept models) with a generalized least squares (GLS) algorithm. These models offer more reliable and efficient estimates than does the two-step regression approach proposed by Jasso (2006; for more details, see the online supplement). Based on the coding of the dependent variable (–5 = wages are "unfairly low"; +5 = "unfairly high"), positive (negative) signs for regression coefficients indicate VPs are more likely assessed to be over- (under-)paid when independent variables increase in value. Thus, positive values of the regression coefficient for gender β_{female} (1 = female VP) indicate ratings that support lower wages for female VPs. At the same time, as the effect size of this regression coefficient increases, the just female-to-male pay ratio that respondents supported declines (i.e., the just gender pay gap grows).

To summarize, we ran regressions that included the vignette dimensions, respondent characteristics (gender, region of residence), and information about the share of men and women in the (former) occupations of respondents and vignette persons, as well as the gender pay ratios in these occupations. We ran models for the pooled data including all respondents and to test for influences of the gender pay ratios experienced on the job for respondents who (formerly) participated in the labor market. In two applications, we estimated the coefficients separately for the experimental groups with more or less

information on the VP and for male and female respondents.

ANALYSES AND RESULTS

Our presentation of the regression results follows the chronology (presented in the summary table) of the assumptions derived from the different theories (Table 1). We report significance levels up to the $p < .10$ level because of the relatively small statistical power of (cross-level) interactions.[11]

Same-Gender Referent Hypothesis

This hypothesis states that female respondents have lower reference standards and hence support lower fair base wages, independent of the gender of the VP, which is particularly true for women working in typical female occupations. Table 3, Model 1, shows the baseline regression model with the effects of the vignette dimensions and some respondent characteristics.

The effect of respondent's gender is of most interest here. The coefficient is close to zero and insignificant, indicating there were no differences in judgments of the average justice perception (fair base wages) between men and women. The second model restricts the sample to respondents who were at least part-time employed (at the time of the survey or formerly), and the third and fourth models also include the gender composition of respondents' occupations as well as the cross-level interaction with respondents' gender. Restricting the analyses to this limited sample, or including the main effect of the gender composition, does not change results (see Models 2 and 3). In Model 4, we see that working in a female-dominated occupation (at least 70 percent women) does not influence respondents' overall judgment of the vignettes: the interaction effect with gender composition is close to zero, not significant, and indicates that the effect is equal for male and female respondents. That is, the average justice evaluation does not change regardless of whether people work in female-dominated

occupations (these results remain stable when using other splits or metric representations of the share of females within occupations).

The effect of the VP's gender will be discussed in more detail in the following section, but so far, we can see that the vignette variables in Table 3 show plausible effects overall. For instance, the impact of VPs' age was both negative and highly significant. In other words, older VPs were more likely than younger VPs to be evaluated as under-rewarded (seniority reward). The same was true for VPs who worked in high-prestige occupations (occupation status reward; see the negative effect of SIOPs) or had a high educational level (educational reward). The high explained variance of approximately 71 percent (see the R^2 values) suggests that fairness evaluations were guided largely by the vignette dimensions, and respondents showed a remarkable amount of social consensus within their evaluations. This is further evidence that male and female respondents have no pronounced differences in their evaluation styles.[12]

Reward Expectations Hypotheses

When directly asked in a single-item question (which was conducted later in the questionnaire after people evaluated the vignettes, results not presented here), nearly all respondents (94.5 percent) stated that an employee's gender should not matter at all for the assessment of fair earnings. Compared to other input factors (i.e., the 11 vignette dimensions in addition to earnings), gender was evaluated as the least important. By contrast, an employee's job performance was evaluated as most important (91.3 percent of our sample evaluated performance as important). Under direct questioning, respondents seemingly shared the norms of equal treatment of male and female employees and performance-based salaries.

However, as regression Table 4 shows, the same respondents' vignette evaluations draw a different picture. The regression coefficient for female VPs is both positive and highly

Table 3. Regression of Justice Evaluations on Vignette Dimensions and Respondent Characteristics (Generalized Least Square Regressions) to Test the Same-Gender Referent Hypotheses 1.1 and 1.2

	(1)	(2)	(3)	(4)
	Full Sample	(Formerly) Employed Resp.	(Formerly) Employed Resp.	(Formerly) Employed Resp.
	β/se	β/se	β/se	β/se
Vignette dimensions				
Gender [1 = female]	.233***	.233***	.233***	.233***
	(.020)	(.026)	(.026)	(.026)
Age	−.007***	−.009***	−.009***	−.009***
	(.001)	(.001)	(.001)	(.001)
Without degree	ref.	ref.	ref.	ref.
Vocational training	−.398***	−.437***	−.437***	−.437***
	(.025)	(.033)	(.033)	(.033)
University degree	−.631***	−.662***	−.662***	−.662***
	(.025)	(.033)	(.033)	(.033)
SIOPS [×10]	−.454***	−.466***	−.466***	−.466***
	(.006)	(.008)	(.008)	(.008)
Gross earnings (ln)	2.873***	2.922***	2.922***	2.922***
	(.011)	(.014)	(.014)	(.014)
Respondent characteristics				
Respondent female	.001	.030	.038	.057
	(.038)	(.048)	(.054)	(.064)
Respondent female-dominated occupation			−.018	.028
			(.054)	(.099)
Interaction				
R_female × R_female-dominated occupation				−.066
				(.118)
Constant	.242***	.331***	.334***	.326***
	(.052)	(.069)	(.069)	(.071)
R-squared	.707	.720	.720	.720
N Vignettes	26,207	15,132	15,132	15,132
N Respondents	1,604	938	938	938
JGPR	.922	.923	.923	.923
	(.006)	(.008)	(.008)	(.008)

Note: R_: respondent. JGPR: just gender pay ratio, calculated as $\exp[\beta\_VP_{female}/ -\beta\_VP_{ln(earnings)}]$. Standard errors are in parentheses.
$+p < .10$; $*p < .05$; $**p < .01$; $***p < .001$ (two-tailed test).

statistically significant ($p < .001$), suggesting that a VP's gender is a meaningful input variable for fairness evaluations. Model 2 tests whether male and female respondents equally contributed to this unequal treatment of male and female rewardees. The interaction effect between respondents' gender and VPs' gender is statistically insignificant. This means male and female respondents assigned lower fair earnings to female employees to the same extent. Figure 2 highlights that both male and female respondents rated male VPs, on average, as underpaid but rated female VPs, on average, as overpaid. Note again that this gender gap cannot be caused by male and female VPs showing different labor market

Table 4. Regression of Justice Evaluations on Vignette Dimensions and Respondent Characteristics (Generalized Least Square Regressions) to Test the Reward Expectations Hypotheses 2.1, 2.2, and 2.3

	(1)	(2)	(3)	(4)
	Full Sample	Full Sample	(Formerly) Employed Resp.	(Formerly) Employed Resp.
	β/se	β/se	β/se	β/se
Vignette dimensions				
Gender [1 = female]	.233***	.207***	.224***	.891***
	(.020)	(.029)	(.026)	(.184)
Age	−.007***	−.007***	−.009***	−.009***
	(.001)	(.001)	(.001)	(.001)
Without degree	ref.	ref.	ref.	ref.
Vocational training	−.398***	−.398***	−.414***	−.412***
	(.025)	(.025)	(.033)	(.033)
University degree	−.631***	−.631***	−.717***	−.714***
	(.025)	(.025)	(.034)	(.034)
SIOPS [×10]	−.454***	−.454***	−.442***	−.443***
	(.006)	(.006)	(.009)	(.009)
Gross earnings (ln)	2.873***	2.873***	2.937***	2.936***
	(.011)	(.011)	(.014)	(.014)
VP_GPR			.340***	.465***
			(.073)	(.098)
VP_male-dom. occ.			−.298***	−.275***
			(.028)	(.040)
Respondent characteristics				
Respondent female	.001	−.025	.029	.030
	(.038)	(.043)	(.049)	(.049)
R_GPR			.013	.268
			(.157)	(.180)
Interactions				
VP_female × R_female		.050		
		(.040)		
VP_female × R_GPR				−.498**
				(.170)
VP_female × VP_GPR				−.262+
				(.135)
VP_female × VP_male-dom. occupation				−.044
				(.054)
Constant	.242***	.255***	.182	−.151
	(.052)	(.053)	(.162)	(.186)
R-squared	.707	.707	.722	.722
N Vignettes	26,207	26,207	15,132	15,132
N Respondents	1,604	1,604	938	938
JGPR	.922	[a]	[a]	[a]
	(.006)			

Note: R_: respondent; VP_: vignette person; GPR: gender pay ratio (mean female-to-male earnings). JGPR: just gender pay ratio, calculated as $\exp[\beta\_VP_{female} / -\beta\_VP_{ln(earnings)}]$. Standard errors are in parentheses.
[a]JGPR not reported because of the inclusion of interaction terms with VP's gender in the regression model.
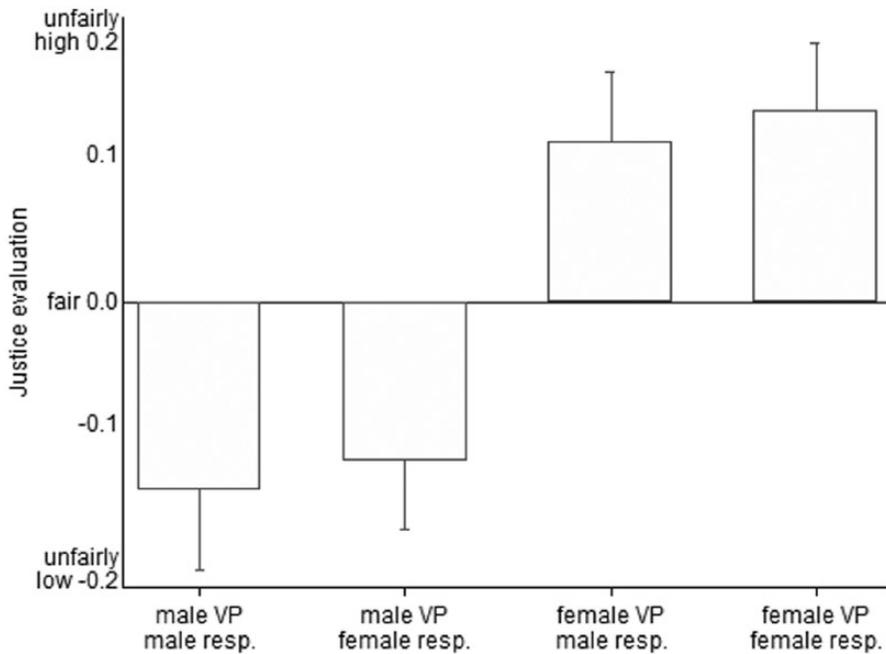+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed test).

Figure 2. Mean Ratings of Male and Female Vignette Persons by Respondents' Gender
Note: The figure shows the mean evaluations of male and female vignette persons given by male and female respondents attached with a standard error. The coefficients result from a generalized least square (GLS) regression controlling for the vignette person's age, educational level, occupation, and gross earnings (adjusted to the mean). $n = 13{,}840$ vignette judgments of $n = 845$ female respondents; and $n = 12{,}367$ vignette judgments of $n = 759$ male respondents.

characteristics: by design, these features were standardized. In other words, we found clear evidence for discrimination against female employees. To determine the size of the gap, we calculated the ratio in terms of gross earnings. The estimated ratio (adjusted for VPs' age, educational degrees, and occupations) is .92 (95% *CI* [.91; .93]): in the eyes of respondents, wages paid to women should be (on average) 92 percent of wages paid to equally qualified men (for the technical details, see the online supplement; further just gender pay ratios [JGPRs] are reported in the bottom lines of our regression tables). The just gender pay ratio—adjusted for labor market characteristics such as education—is remarkable and is an amount comparable to the actual adjusted ratio in Germany (which is approximately .88; Gartner and Hinz 2009).

To investigate the mechanisms more deeply that lead to this just gender pay ratio, Models 3 and 4 include the gender pay ratios

of the occupation in which the respondent is working and the gender pay ratio of the VP's occupation. Model 4 estimates cross-level interactions with the actual gender pay ratios experienced by respondents in their occupations. As mentioned in the Methods section, we imputed these gender pay ratios using calculations made with a large-scale population survey (SOEP). The interaction effect is negative and significant, meaning that respondents who worked in an occupation with high gender pay ratios (low gender inequalities) also justified higher just gender pay ratios (lower inequalities) in their evaluations. This is in line with the first specific prediction derived from reward expectations theory. Thus, our data suggest that just gender pay ratios are framed predominantly by the gender pay ratios that respondents experience in their own occupations; however, they are also framed by the gender pay ratios in the occupations they are required to evaluate.
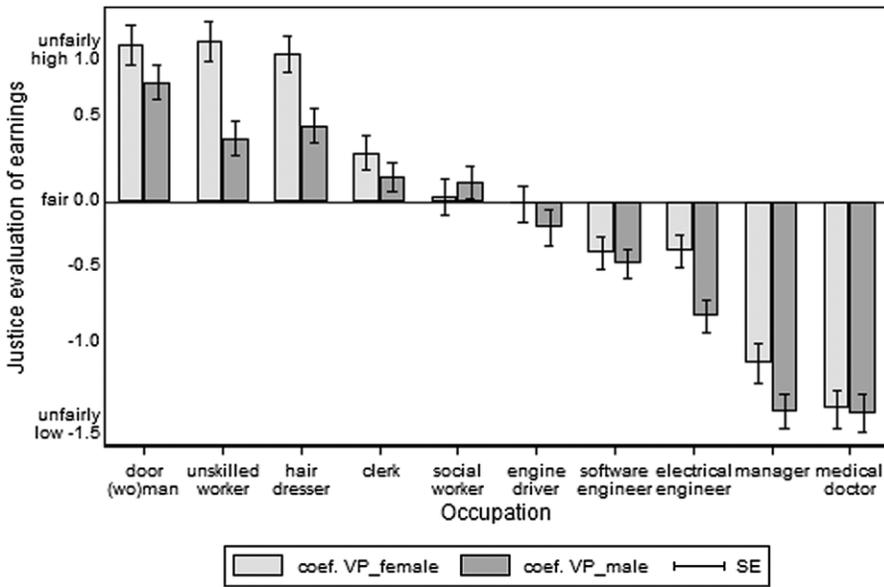
Figure 3. Gender Discrimination by Occupation
Note: The figure shows regression coefficients for the 10 occupations for male and female vignette persons (VPs) separately attached with a standard error (SE). The coefficients result from a regression that controls for the other vignette dimensions of age, education, and gross earnings (adjusted to the mean). $n = 13,840$ vignette judgments of $n = 845$ female respondents; and $n = 12,367$ vignette judgments of $n = 759$ male respondents.

With respect to theoretical approaches, these results provide strong evidence for reward expectations theory. In addition, respondents in West Germany tended to support lower fair pay ratios (i.e., larger gender inequalities) than did respondents in East Germany. The difference (again tested by an interaction term) did not reach statistical significance, which might be caused by its low statistical power, as only approximately one-fifth of respondents lived in East Germany (results not reported, but available on request).

Our respondents did not support particularly small just pay ratios (i.e., large inequalities) for male occupations—the interaction with the VP working in a male-dominated occupation is insignificant (see Model 4 in Table 4). Figure 3 provides a detailed representation of gender differences for all occupations. Social work is the only occupation in which female employees were assigned comparatively higher earnings (in Germany, social work is typically a female occupation). Moreover, there was marked discrimination against female VPs working in the typically

male profession of management. Both of these results are in line with reward expectations theory. What is striking, however, is the similarly large discrimination observed for gender-neutral occupations, such as clerks.

Double Standards

Model 1 in Table 5 tests the hypothesis, derived from the extension of reward expectations theories in the form of the double standard, that performance is interpreted in a gender-specific way. To conduct this test, we added two interaction terms between the VP's gender and job performance (above/below average, included in the 12-dimension split only). The reference category is average performance. Both interaction terms are statistically significant, meaning that information about performance is indeed evaluated differently for male and female VPs. In other words, double standards seem to exist with respect to performance evaluations, but the effects are not exactly as expected. Double standard theory predicts high performance

Table 5. Regression of Justice Evaluations on Vignette Dimensions (Generalized Least Square Regressions) to Test the Reward Expectations Hypothesis 2.4 and Statistical Discrimination Hypothesis 3

	(1)	(2)	(3)	(4)
	12 Dimensions	5 Dimensions	8 Dimensions	12 Dimensions
	β/se	β/se	β/se	β/se
Vignette dimensions				
Gender [1 = female]	.270***	.233***	.194***	.272***
	(.061)	(.035)	(.034)	(.035)
Age	−.005***	−.009***	−.005***	−.005***
	(.002)	(.002)	(.001)	(.002)
Without degree	ref.	ref.	ref.	ref.
Vocational training	−.277***	−.586***	−.341***	−.274***
	(.043)	(.043)	(.042)	(.043)
University degree	−.465***	−.800***	−.605***	−.465***
	(.044)	(.044)	(.043)	(.044)
SIOPS [×10]	−.373***	−.553***	−.444***	−.373***
	(.011)	(.011)	(.011)	(.011)
Gross earnings (ln)	2.824***	2.907***	2.903***	2.826***
	(.019)	(.019)	(.018)	(.019)
Experience [1 = much]	−.403***		−.427***	−.384***
	(.046)		(.045)	(.046)
Job tenure [1 = long]	.021		−.171***	.009
	(.042)		(.041)	(.042)
No. of children	−.088***		−.105***	−.089***
	(.012)		(.011)	(.012)
Performance: below average	.381***			.285***
	(.061)			(.043)
Performance: average	ref.			ref.
Performance: above average	−.385***			−.298***
	(.061)			(.042)
Health status [1 = health problems for a long time]	−.015			−.007
	(.037)			(.037)
Organization size: small	ref.			ref.
Organization size: medium	−.058			−.072+
	(.043)			(.043)
Organization size: large	−.058			−.057
	(.042)			(.042)
Economic situation: threatened by bankruptcy	ref.			ref.
Economic situation: solid	−.214***			−.211***
	(.042)			(.042)
Economic situation: high profits	−.275***			−.272***
	(.042)			(.042)
VP_female × below average perf.	−.191*			
	(.088)			
VP_female × above average perf.	.171+			
	(.087)			

(continued)

Table 5. *(continued)*

	(1)	(2)	(3)	(4)
	12 Dimensions	5 Dimensions	8 Dimensions	12 Dimensions
	β/se	β/se	β/se	β/se
Constant	.787***	.379***	.634***	.780***
	(.101)	(.083)	(.087)	(.098)
R-squared	.714	.713	.722	.713
N Vignettes	8,792	8,698	8,717	8,792
N Respondents	545	528	531	545
JGPR	[a]	.923	.935	.908
		(.011)	(.010)	(.011)

Note: VP_: vignette person. JGPR: just gender pay ratio, calculated as $\exp[\beta\_VP_{female}/ -\beta\_VP_{ln(earnings)}]$. Standard errors are in parentheses.
[a]JGPR not reported because of the inclusion of interaction terms with VP's gender in the regression model.
$+p < .10$; $*p < .05$; $**p < .01$; $***p < .001$ (two-tailed test).

will more likely be devalued for women (i.e., less strongly rewarded in the form of higher fair pay), whereas women's low performance will more likely be recognized (i.e., penalized in the form of lower pay). Although the former tended to be true, low performance was less strongly considered for female than for male VPs (both interaction effects lower the respective main effects of low and high performance). In general (i.e., regardless of VPs' performance level), respondents were more sensitive to the performance of male than female VPs.

Add-Ons: Statistical Discrimination and the Differential Norm Hypothesis

We now turn to our final body of hypotheses derived from statistical discrimination and the differential norm hypothesis. Models 2, 3, and 4, presented in Table 5, focus on statistical discrimination by measuring the influence of additional information about employees. Recall that in different experimental splits, respondents evaluated vignettes with differing amounts of information on employees' characteristics (5, 8, or 12 vignette dimensions).

Model 2 shows results for respondents who evaluated vignettes with only five dimensions.

The effect for female VPs is positive and statistically significant, indicating a just gender pay ratio disadvantaging women. Model 3 provides results for respondents who evaluated the split with eight dimensions. Within this split, we added additional dimensions on VPs' labor market experience, job tenure, and number of children. The effect of gender is remarkably robust in regard to this additional information. This is also true for Model 4, which presents the fairness evaluations of respondents who rated vignettes in light of even more information about VPs' characteristics (12 vignette dimensions). In addition to extensive information, now including additional information about an organization's size and economic situation, as well as VPs' health status and job performance, the coefficient hardly changes in effect size, continuing to show a highly significant, positive effect. This is clear evidence against the statistical discrimination theory, which states that gender serves only as a proxy for missing information on performance or labor market experience. Statistical tests confirmed that the gender coefficient did not differ across the three experimental splits.[13]

Another interesting observation when comparing Model 2 with Models 3 and 4 is
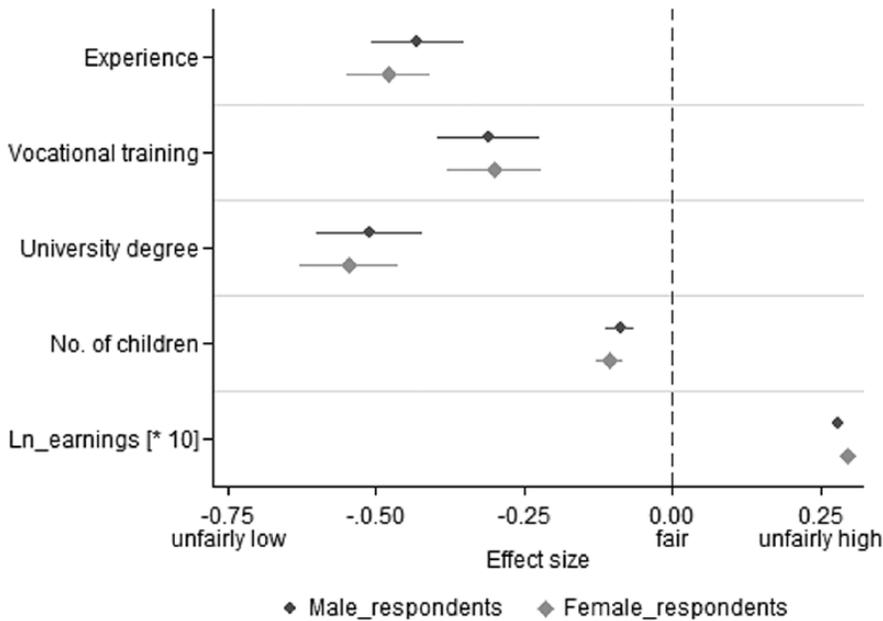
Figure 4. Application of Justice Dimensions by Gender of Respondents
Note: The figure shows regression coefficients of the vignette dimensions related to equity and need by male and female respondents attached with a standard error. The coefficients result from a generalized least square (GLS) regression with $n = 9,284$ vignette judgments of $n = 569$ female respondents and $n = 8,225$ vignette judgments of $n = 507$ male respondents. Significance tests were applied in pooled interaction models. We found no significant differences between men and women in regard to experience ($\chi^2 = 1.07$; $p = .300$), vocational training ($\chi^2 = .03$; $p = .868$), university degree ($\chi^2 = .15$ $p = .696$), and number of children ($\chi^2 = 1.18$; $p = .278$). The coefficient of ln_earnings is significantly higher for female respondents than for male respondents ($\chi^2 = 46.51$; $p < .001$). Part E of the online supplement contains the regressions table.

that the impact of VPs' age declined, which is in line with the theory of statistical discrimination. The age effects in Models 3 and 4, with additional information such as experience and performance added, are around half the size of the effect estimated in Model 2, where only some basic information on the VP was provided. When respondents had no direct information about labor market experience or seniority, they tended to use age as a proxy (labor market economists frequently regard age as a proxy for missing information on labor market experience; see, e.g., Mincer 1974). Gender, however, is a different case: the observed pattern fits the prediction that gender beliefs resist information that invalidates stereotypes about women showing different performance than men.

The differential norm hypothesis predicts that female and male respondents have different justice norms and therefore evaluate different input factors, such as education or number of children, differently. These differences are deemed to be related to differences in salient or dominating justice principles, such as equity or need. Figure 4 shows the regression coefficients for the vignette dimensions connected to equity (experience and education [vocational training and university degree; without vocational training is the reference category]), need (number of children), and earnings. We found no differences in the fair rates of return on vocational training, university degree, experience, or children between male and female respondents (see Part E of the online supplement for the interaction model). Thus, respondents seem to apply the same norms when evaluating justice and particularly focus on the dimensions related to the equity principle, which supports

prior results reported in the literature (Kluegel and Smith 1986).

Validity and Robustness Checks

The principal advantage of the FS experiment is that it offers the possibility of studying the evaluation of male and female employees who are described by the same levels of labor market endowment, thereby promising mostly unbiased assessments of gender discrimination (high internal validity). At the same time, the broad population sample enables the study of the framing effects caused by respondents' workplace experience; it also provides a high degree of generalizability (external validity). However, numerous potential threats to validity remain.

First, one must ensure that all the vignette stimuli and the three experimental splits on amount of information (5, 8, or 12 vignette dimensions) are randomly assigned to sample members. We address this potential problem by checking whether our realized sample continues to constitute a valid random experiment; to perform this check, we must confirm that the factors defining each vignette (i.e., the experimental stimuli) are uncorrelated with both observed and unobserved individual characteristics. It is straightforward to test for correlation with observed characteristics. We calculated correlation coefficients between the vignette dimensions and six respondents' characteristics (age, gender, occupational status, education, vocational training, East/West Germany). These correlation coefficients are all below $|r| = .01$. We tested for correlation with unobserved characteristics by estimating fixed- and random-effects regressions with the five core vignette dimensions as predictors, and by testing whether there are statistically significant differences between the two sets of estimates using a Hausman-type test that is robust to potential correlation in the errors within individuals, as suggested by Wooldridge (2010). In effect, this constitutes a test of whether controls in the model are uncorrelated with individual-level unobservable characteristics.

The computed *p*-value of this test is .325. We therefore do not reject the null hypothesis that the coefficients from fixed- and random-effects models are equal, and we accept the hypothesis that the factors are indeed randomly assigned across respondents.

Our estimates may be biased if reported measures of justice are generally not comparable across respondents because they are, for instance, influenced by respondents' state of mind at the time of the interview. This problem is also addressed by the comparison of fixed- and random-effects coefficients. Fixed-effects estimates net out all between-individual differences; the fact that there are no statistically significant differences between fixed- and random-effects analyses demonstrates that, to the extent that unobserved differences in how respondents report feelings of (in)justice exist, those differences cannot bias our results because they are uncorrelated with our vignette factors. In real labor market data, of course, factors describing employees are not generally uncorrelated with individual characteristics of observers, and this can lead to biased estimates.

A further threat to validity exists if respondents cannot cope with the complexity of the experiment. The simultaneous evaluation of 12 dimensions is a relatively complex task that might have motivated respondents to use heuristics, such as fading out some dimensions. This relates to the result that more information (on performance) did not cause the gender effect to vanish. We conducted thorough analyses to test for any signs of cognitive overload (documented in Sauer et al. 2011). We found small indications of cognitive overburdening only for respondents older than 60 years of age and respondents not included in the labor market. Excluding those respondents did not change the results. Moreover, the dimension "performance of VP"—a key variable in the hypothesis tests—shows a very strong impact on evaluations throughout the models, which is likely the strongest proof that respondents noticed this piece of information. Similarly, the observation that the effect of gender becomes somewhat stronger in the

12-dimension condition is evidence against the possibility that respondents ignored core vignette dimensions due to information overload.

As in any survey study, social desirability can result in biased estimates. We note that the just gender pay gaps might have been underestimated if respondents were aware of and attempted to mask their discriminatory attitudes. However, the FS task, with its simultaneous evaluation of multiple dimensions, is considered more likely than alternative methods (including randomized response techniques) to be immune to social desirability bias (Auspurg and Hinz 2015). Additionally, even if the gender effect was somewhat downward biased due to social desirability, it remained sufficiently strong to reveal a meaningful causal relationship.

Furthermore, because we used adjectives and not exact numerical levels to describe vignette dimensions such as work experience, even when providing a lot of information on the VP in the 12-dimension split, these dimensions might have been interpreted in a gender-specific way, hence undermining the intended standardization of information in the experiment. In that case, one would expect idiosyncrasies particularly in the interpretation of information about female VPs, because women's labor market careers are, in general, less standardized than those of men, and some (but not all) respondents might have interpreted "much" or "low" experience in relation to typical female and male careers. However, we found hardly any signs that unobserved heterogeneity (measured as variance of error terms) was higher for the evaluations of female VPs, nor did we find that effect sizes of the labor market experience dimension showed a higher variation for descriptions of older female VPs (50 years or older compared to 30 or 40 years). This piece of information, in particular, might have caused uncertainties among respondents as to what the qualitative levels of labor market experience stood for.[14]

Respondents' use of response scales in FSs was recently the subject of lively debate. The indirect rating task routinely used in FS experiments (rather than directly asking respondents about fair wage amounts) was criticized for possible anchor effects, that is, respondents might adapt their ideas of fair pay to the pay schemes presented in vignettes (Markovsky and Eriksson 2012; for a rejoinder, see Jasso 2012). However, in our vignette universe (in contrast to real labor markets), the earnings and gender of the vignette persons were uncorrelated ($r = .01$). Thus, if there were strong anchor effects, we would not have observed any just gender pay ratios. Finally, a serious threat to construct validity exists when respondents evaluate the plausibility of the vignettes rather than justice (Faia 1980). In that case, the observed correspondence between just and actual pay ratios is merely a methodological artifact. Unfortunately, there has been no research on this issue. We therefore decided to run a small pilot study as an add-on to our substantive main experiment. Approximately 90 university students were asked to evaluate the same vignettes used in this main survey; one half of the students were asked to rate the fairness, and the other half rated the plausibility of the vignette cases (for details, see the online supplement). The results clearly suggest that respondents differentiate between fairness and plausibility ratings, thus confirming the ability of FS studies to assess *fairness* attitudes. In addition, these experiments not only demonstrated an appropriate level of public knowledge about actual wage schemes, but they also provided further evidence for the priming processes hypothesized by justice theories (see again the online supplement).

SUMMARY

We summarize the results of our analyses in five points. First, we found no difference in the overall justice judgments by respondents' gender and therefore no hints of diverging comparison standards. Women do not generally have lower expectations of fair earnings than men. The same-gender referents hypothesis is not supported by our data.

Second, the vignette dimension gender had a strong impact on the vignette judgments for both male and female respondents favoring male vignette persons. The just gender pay gap found in earlier studies (Jasso and Webster 1997) is supported by both sexes. The effect was stronger for respondents who experienced gender differences in pay in their own occupations. Moreover, the gender pay ratio in the vignette person's occupation played a role. Therefore, experienced inequalities affected justice perceptions and ratings. Reward expectations, and therefore justice evaluations, appear to be learned in social contexts that are more or less gender unequal. This inequality is internalized and deemed to be fair. Note that this result could not be detected with standard item measurements with the same respondents. One reason might be social desirability bias—people attempt to stick to a norm they learned to be desirable but not to the norm they have internalized. Another reason is that gender *unconsciously* cues individuals' reward expectations, which would also be in line with reward expectations theories (Correll and Ridgeway 2003; Thébaud 2015). To find these deeply rooted norms, the factorial survey offered a useful tool and clearly showed that status differences between men and women are still salient and work strongly. The only finding that did not conform to reward expectations theories was that just gender pay ratios did not intensify for typical male occupations; the evidence regarding gender type of occupations was mixed overall. Nevertheless, by and large, our results provide strong support for reward expectations theories.

Third, additional information about the person described in the vignettes provided no change in the gender bias, which clearly indicates that we do not have an informational bias (as the statistical discrimination argument would assume), but rather a cognitive bias, as status theories assume (Ridgeway 2011). This finding clearly supports the assumptions of subtle mechanisms of gender discrimination and self-fulfilling prophecies developed by sociologists (Quillian 2006): long-lasting

differences in (fair) earnings do not necessarily need to match differences in true performance to be legitimized; gender has a status value on its own and is not only a proxy for unmeasured skills or competencies.

Fourth, we find some evidence of double standards regarding performance on the job. This dimension seemed to be less important in the evaluation of female than in male vignette persons. These results are not exactly in line with findings of other researchers who investigated gendered double standards. Whereas previous research has found that women's *high* performance is devalued, our findings suggest that both high- and low-performance information is more often ignored for female employees; in other words, judgments of male vignette persons depend more on performance criteria than do judgments of female vignette persons.[15]

Fifth, there seem to be no general differences in the justice norms men and women apply when judging the fairness of earnings. Men and women rated the dimensions connected to equity and need very similarly, focusing mainly on the equity principle. Thus, not only the absolute importance of equity and need in justice judgments, but also the extent to which men and women weight these reward-relevant characteristics in relation to each other are very similar.

DISCUSSION

We began our analysis with the awareness that women continue to earn remarkably less than similarly qualified men, and there is strong evidence that this and other (gender) inequalities are at least partially caused by discrimination (Abbott 2005). The persistence of discrimination in Western societies is puzzling in light of not only the many policy measures against it (e.g., legislation strictly forbidding it) but also the equality norms held by most individuals. Our research supports the idea that gender discrimination likely survives despite policy prescriptions and market forces because it is also tolerated by the disadvantaged group, that is, women. The

gender ratio found in our study, with fair earnings for women at about 92 percent of that of similarly equipped men, might be called modest in effect size, but it was remarkably stable in all model specifications. Additionally, repeated over several occasions and contexts during a career, even modest effects can accumulate to large gender inequalities (Eagly and Karau 2002; Ridgeway 2001; for a computer simulation, see Martell, Lane, and Emrich 1996). Our research on Germany confirmed findings for the United States (Jasso and Webster 1997), which shows that both men *and* women consider comparatively lower wages for female employees to be fair even if these employees are characterized by identical labor market attributes.

Thus, researchers who are interested in explaining gender discrimination must address the following question: what makes members of a disadvantaged group, such as women or ethnic minorities, participate in discrimination? In this article, we analyzed two different theoretical approaches: same-gender referents (i.e., women compare themselves mostly with women) and reward expectations theory (i.e., gender works as a general status variable). Our results clearly contradict the same-gender referents hypotheses while supporting reward expectations theory. In general, men and women do not have different reference standards when evaluating pay gaps, but men's *and* women's expectations about just earnings for both sexes are imprinted by specific experiences in the labor market: that is, the social construction of status mediated by social interactions in the workplace, where group categories such as gender are connected to status beliefs in higher performance and fair rewards for one status group (Berger and Fişek 2006; Ridgeway 1997). The extension of double standards theory posits that performance is perceived in a gender-colored, stereotype-confirming manner (Foschi 1996, 2000), which would make gender status beliefs even more resistant to change. We found some evidence that for male employees, performance is crucial for the evaluation of earnings, whereas respondents hardly acknowledged the performance levels of female employees.

In addition, we can rule out that statistical discrimination and gender-specific applications of justice norms (need versus equity) are important driving forces to legitimate gender pay gaps. Gender seems more powerful as an organizing principle of social interactions and inequality than as a simple proxy for missing information. This finding shows the surplus of sociological theories that go beyond economic approaches to discrimination (for these theories, see Fang and Moro 2011). Because men and women support gender differences to the same extent, these views might not be explained simply by self-interest and "rational" in-group favoritism. This deeply internalized acceptance of gender differences and inequality likely makes gender status beliefs strong, self-fulfilling, and thus stable over a long period of time.

Although gender status beliefs might not be very "strong" (the effect size in our study was "modest"), they are ubiquitous and may coexist with egalitarian attitudes. Performance expectations, for example, regarding educational investments and work effort, likely function as self-fulfilling prophecies. These expectations might trigger even greater gender inequalities, either because men are boosted by higher performance expectations or because women who internalize status beliefs of lower competence and status exhibit lower work commitment. In addition, social psychology work shows that *stereotype threats* trigger anxiety that fulfills the stereotypes of lower performance (Steele and Aronson 1995).

For all these reasons, the inequalities fueled by status beliefs might accumulate and reach a remarkable ability to resist social change (Ridgeway and Correll 2004). On the one hand, justice attitudes affect individuals' labor market behavior (e.g., their work performance); on the other hand, these social practices shape justice attitudes. These two mechanisms are not just two sides of the same coin, they are intertwined: inequalities at the

macro-level frame the belief systems and reward expectations of individuals interacting at the micro-level of organizations and the workplace. Aggregating these day-to-day interactions confirms inequalities at the macro-level of societies.

With respect to policy interventions, our results suggest that single pieces of disconfirming information are insufficient to stop discrimination. The "stall in the gender revolution" will certainly not resolve itself. Rather, more transparency about pay gaps might even increase the internalization of gender status beliefs.[16] According to the empirically supported reward expectations theory, frequent interaction with permanent female role models in high-status, high-resources positions is needed to reduce gender inequalities. Such role models will only appear, however, if policy measures provide organizations with effective incentives to hire more women in leading positions. At the same time, reward expectations theory suggests a more nuanced understanding of the unintended consequences of mother-friendly social policies: if measures such as long parental leaves are not equally taken by both sexes, they tend to confirm ideas about women's lower employment commitment, and hence status. Paradoxically, such policies might encourage employer discrimination (for evidence, see Mandel and Semyonov 2005).

With respect to an adequate theoretical understanding of gender pay ratios, reward expectations theory might interact with processes of statistical discrimination on the employer side, which was not the focus of this study (for similar arguments, see Correll and Benard 2006; Correll et al. 2007). Status beliefs can create a strong basis for true productivity differences. Simultaneously, employers react to these differences and contribute to the persistence of gender inequalities by filtering men and women into status hierarchies, thus reinforcing the conditions that trigger the construction of status beliefs. The results of our study underline that it is crucial to look beyond (true) differences in performance, and hence arguments that are often the most

terminal by economists (for a recent review, see Fang and Moro 2011). The interplay of same-gender referent effects and status beliefs is less clear. Status beliefs motivate the choice of same-status comparison with others, but in particular, comparisons *across* status categories intensify the salience of status beliefs. This might be one reason why we found support for only one of the two theories, although these theories are based on similar processes of social comparisons.

Our study does have some limitations. If one wants to explain gender pay ratios, the employer perspective is also relevant. Our research used evaluations of respondents from a general population survey—very few of whom were in the role of employers. Thus, an FS study on hiring behavior could be informative in completing the picture. In particular, more research on the link between employers' discriminatory attitudes (or justice principles) and their actual behavior would be desirable. Expectation states theories offer particularly promising tools to study the *situational mechanisms*, that is, the mechanisms of how current distributions of rewards frame justice beliefs; future research should focus on the *action-formation mechanisms*, that is, how justice attitudes affect social behavior (see also Jasso and Rossi 1977; for a typology of different mechanisms, see Hedström and Swedberg 1998).[17]

Furthermore, research on the priming of status beliefs could be strengthened through the collection of additional information about people's actual labor market conditions and possible other contexts in which priming processes take place. We found that regional contexts tend to matter (i.e., East versus West Germany), but we did not have enough power (number of observations) or information on the length of exposure (moving behavior between regional contexts) to investigate these effects in detail. Equity and status value theories are vague with regard to which regional or socialization contexts matter. Not least because of the finding from prior research that university students do not support just gender pay gaps, future research

should conduct longitudinal studies on belief-formation over the life course.

We mainly used categorical dimensions to present male and female vignette persons. In future applications, one might replace some of these categories with quantitative indicators to provide even more confidence that respondents understand that male and female vignette persons (on average) show the same (performance) levels. In our application, proxies such as metric test scores would not have fit all vignette cases well, but one might draft scenarios where this is more apt; more research on double standards is needed. With regard to the selection of occupations of vignette persons, future studies might test different specifications of occupational groups (e.g., micro classes) instead of the (ISCO) occupations we used. Broader occupational categories might define relevant frames for status beliefs while at the same time allowing somewhat more reliable calculations of gender pay ratios based on higher numbers of cases.

Future studies might ideally be implemented in organizational research using a longitudinal design in which thorough information about gender distributions within working groups, status positions, and (fairness) attitudes is collected, while variation in regional contexts is simultaneously included (Baron and Newman 1990; Tomaskovic-Devey and Skaggs 1999). Linking the real-world macro-context of societies more systematically to the micro-mechanisms where the priming of status beliefs and other stereotypes takes place would enhance our knowledge not only of the roots of gender pay but also of other social inequalities.

Acknowledgments

Funding

Notes

1. Research on pay referents suggests that individuals are most likely to compare themselves with others who work in the same occupation. Prior work finds "market referents" are more important than local comparisons with co-workers or social referents such as family members, friends, and relatives (Bygren 2004).

2. These settings consist of procedures to manipulate key independent variables (e.g., status characteristics) and observe their effects on the dependent variable, which is typically a measurement of social influence. A common task is "contrast sensitivity," an ability introduced as being a newly discovered skill (Berger et al. 2014; Kalkhoff and Thye 2006). Subjects are asked to judge slides containing black-and-white rectangles in a checkerboard style and to indicate which of the two colors dominates. The task is actually ambiguous (both colors are equally present) to allow for measurement of the acceptance and rejection of social influence. The extent to which participants are influenced by other participants is used as a clue for performance expectations and status beliefs.

3. In addition, several analyses are based on lab experiments or labor market data (e.g., Altonji and Pierret 2001; Castillo and Petrie 2010). These analyses often demonstrate that the reasoning underlying statistical discrimination is viable, at least in the laboratory. It remains difficult to demonstrate this form of discrimination in real markets, however, as there are no measures of marginal productivity. Therefore, the theory's main assumption—that discrimination is based only on characteristics predictive of productivity—cannot be tested (Arrow 1998).

4. Vignettes were set up with the gender of the vignette person being uncorrelated (orthogonal) to all other vignette dimensions. In the realized sample, the random allocation of vignettes to respondents or (unit) non-response might nevertheless trigger some correlations. We therefore also carefully checked the confounding patterns in the realized sample. The maximum correlation of one of the 12 vignette dimensions with gender of the vignette person was $r = .06$. In addition, we did not find any substantial correlation between a vignette dimension (including vignette person's gender) and a respondent's characteristics (see the section on "robustness checks" for further details).

5. The splits with 10, 20, and 30 vignettes and 5, 8, and 12 dimensions were completely crossed with

each other. Each of the resulting nine experimental cells had at least 2,719 valid vignette judgments.

6. For this survey, a Resolution IV-design (Auspurg and Hinz 2015) was realized, that is, some—but not all—two-way interactions between vignette dimensions were orthogonalized.

7. The vignettes do not provide information about marital status. In another study, we included this information and found no gender differences in the evaluation; we also found no differing patterns in the evaluation of the dimension "number of children." Results of this study, which was part of a larger pre-test ($n = 1,066$), can be found in Sauer and colleagues (2014).

8. For instance, in the year of our survey, 83.4 (37.0) percent of employed part-time (full-time) workers who contributed to social security insurance were women (December 2009; authors' calculation based on Bundesagentur für Arbeit 2014).

9. The German SOEP is equivalent to the Panel Study of Income Dynamics (PSID) in the United States.

10. We estimated the following regression models: $J_{ij} = \alpha + \beta_{female}R\_female_j + \beta_1 VP\_female_i + \beta_2 VP\_age_i + \beta_3 VP\_vocational_i + \beta_4 VP\_university_i + \beta_5 VP\_siops_i + \beta_6 VP\_ln(A)_i + u_j + \varepsilon_{ij}$, with R as an indicator for respondents, VP for vignette persons' dimensions, and u_j and ε_{ij} being error terms of level two (respondents) and level one (vignettes), respectively. In some models, we added further respondent or vignette characteristics. We also ran models using random-slopes regressions and—to account for measurement of the dependent variable "justice evaluation" J_{ij}—multilevel ordered logit models. All results remain robust; therefore, we use the linear random-intercept model for ease of interpretation.

11. Following Cohen (1969:367–69), "[a]lthough generally overlooked by behavioral scientists, the power of tests of interactions in a factorial design is distinctly lower than that of the main effects; . . . to avoid doing F tests on interactions which are of low power and hence yield ambiguous negative results, one should consider setting less stringent significance criteria for interaction tests, e.g., $\alpha = .10$ instead of .05. It may well be worth a larger Type I error risk to bring Type II error down to a tolerable level."

12. Additional controls for respondents' age, education, vocational training, employment status, and region of residence did not change the substantive results. Given the experimental design, we used parsimonious models restricted mostly to experimental stimuli.

13. When estimating a pooled regression model with all experimental splits, including interaction effects between VPs' gender and experimental splits, we found the following results: $\beta[8\_dimensions \times VP\_female] = -.027$, $p = .590$; $\beta[12\_dimensions \times VP\_female] = .063$, $p = .204$.

14. Nevertheless, one should replicate the experiments using different operationalizations. Following recent discussions in the literature on discrimination research, there might always be some unobserved heterogeneity left that causes individuals to interpret rewardees' characteristics in a group-specific way (Neumark 2012).

15. We are grateful to one anonymous reviewer for pointing to a possible explanation for this unexpected finding. Respondents might adjust to the long historical tradition of women being systematically disadvantaged, which makes them less attentive to specific information on women's performance. Further research is needed to test this assumption.

16. Policy initiatives in Europe currently aim at implementing such higher transparency. For instance, the Austrian Equal Treatment Act 2011 legally requires all employers to state the (collective) minimum wage in job vacancy advertisements and to produce income reports if the company has at least 150 employees (European Commission 2015).

17. One theoretical starting point could be the "theory of reasoned action" proposed by Fishbein and Ajzen (2010). In brief, this model suggests that behavioral intentions are the main predictor of actions other than behavioral constraints (e.g., missing resources); moreover, these intentions are influenced by normative beliefs, behavioral beliefs, and control beliefs (Fishbein and Ajzen 2010). Normative beliefs are covered by the justice evaluations we studied. Behavioral beliefs point to the negative or positive consequences that individuals expect if they engage in a particular behavior.

References

Abbott, Andrew. 2005. "Sociology of Work and Occupations." Pp. 307–30 in *The Handbook of Economic Sociology*, edited by N. J. Smelser and R. Swedberg. Princeton, NJ: Princeton University Press.

Adams, John Stacy. 1965. "Inequity in Social Exchange." Pp. 267–99 in *Advances in Experimental Social Psychology*, Vol. 2, edited by L. Berkowitz. New York: Academic Press.

Aisenbrey, Silke, Marie Evertsson, and Daniela Grunow. 2009. "Is There a Career Penalty for Mothers' Time Out? A Comparison of Germany, Sweden and the United States." *Social Forces* 88(2):573–605.

Akchurin, Maria, and Cheol-Sung Lee. 2013. "Pathways to Empowerment: Repertoires of Women's Activism and Gender Earnings Equality." *American Sociological Review* 78(4):679–701.

Akerlof, George A., and Janet L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics* 105(2):255–83.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116(1):313–50.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." Pp. 3–33 in *Discrimination in Labor Markets*, edited by O. Ashenfelter and A. Rees. Princeton, NJ: Princeton University Press.

Arrow, Kenneth J. 1998. "What Has Economics to Say about Racial Discrimination?" *Journal of Economic Perspectives* 12(2):91–100.

Auspurg, Katrin, and Thomas Hinz. 2015. *Factorial Survey Experiments*. Thousand Oaks, CA: Sage.

Auspurg, Katrin, Thomas Hinz, and Stefan Liebig. 2009. "Complexity, Learning Effects, and Plausibility of Vignettes in Factorial Surveys." Presented at the 104th Annual Meeting of the American Sociological Association (ASA), San Francisco, CA.

Baron, James N., and Andrew E. Newman. 1990. "For What It's Worth: Organizations, Occupations, and the Value of Work Done by Women and Nonwhites." *American Sociological Review* 55(2):155–75.

Becker, Gary S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* 3(1):S33–S58.

Berger, Joseph, and M. Hamit Fişek. 2006. "Diffuse Status Characteristics and the Spread of Status Value: A Formal Theory." *American Journal of Sociology* 111(4):1038–79.

Berger, Joseph, M. Hamit Fişek, Robert Z. Norman, and David G. Wagner. 1985. "Formation of Reward Expectations in Status Situations." Pp. 215–61 in *Status, Rewards, and Influence: How Expectations Organize Behavior*, edited by J. Berger and M. J. Zelditch. San Francisco, CA: Jossey-Bass.

Berger, Joseph, M. Hamit Fişek, Robert Z. Norman, and Morris Zelditch Jr. 1977. *Status Characteristics and Social Interaction: An Expectation-States Approach*. New York: Elsevier.

Berger, Joseph, David G. Wagner, and Murray Webster Jr. 2014. "Expectation States Theory: Growth, Opportunities and Challenges." *Advances in Group Processes* 31:19–55.

Berger, Joseph, and Murray Webster Jr. 2006. "Expectations, Status, and Behavior." Pp. 268–300 in *Contemporary Social Psychological Theories*, edited by P. J. Burke. Stanford, CA: Stanford University Press.

Bielby, William T., and James N. Baron. 1986. "Men and Women at Work: Sex Segregation and Statistical Discrimination." *American Journal of Sociology* 91(4):759–99.

Blau, Francine D., and Lawrence M. Kahn. 2007. "The Gender Pay Gap." *The Economists' Voice* 4(4):1–6.

Bundesagentur für Arbeit. 2014. "Beschäftigungsstatistik. Sozialversicherungspflichtig Beschäftigte nach ausgewählten Merkmalen." In *Arbeitsmarkt in Zahlen*, edited by Bundesagentur für Arbeit. Nuremberg. Retrieved March 23, 2015 (https://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Themen/Beschaeftigung/Beschaeftigung-Nav.html).

Bygren, Magnus. 2004. "Pay Reference Standards and Pay Satisfaction: What Do Workers Evaluate Their Pay Against?" *Social Science Research* 33(2):206–24.

Bylsma, Wayne H., and Brenda Major. 1992. "Two Routes to Eliminating Gender Differences in Personal Entitlement: Social Comparisons and Performance Evaluations." *Psychology of Women Quarterly* 16(2):193–200.

Bylsma, Wayne H., and Brenda Major. 1994. "Social Comparisons and Contentment: Exploring the Psychological Costs of the Gender Wage Gap." *Psychology of Women Quarterly* 18(2):241–49.

Castillo, Marco, and Ragan Petrie. 2010. "Discrimination in the Lab: Does Information Trump Appearance?" *Games and Economic Behavior* 68(1):50–59.

Charles, Maria, and David B. Grusky. 2004. *Occupational Ghettos: The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.

Cohen, Jacob. 1969. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Colquitt, Jason A., Donald E. Conlon, Michael J. Wesson, Christopher O. L. H. Porter, and K. Yee Ng. 2001. "Justice at the Millennium: A Meta-analytic Review of 25 Years of Organizational Justice Research." *Journal of Applied Psychology* 86(3):425–45.

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., Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *American Journal of Sociology* 112(5): 1297–1338.

Correll, Shelley J., and Cecilia L. Ridgeway. 2003. "Expectation States Theory." Pp. 29–51 in *Handbook of Social Psychology*, edited by J. Delamater. New York: Kluwer Academic/Plenum Publishers.

Davison, H. Kristl. 2014. "The Paradox of the Contented Female Worker: Why Are Women Satisfied with Lower Pay?" *Employee Responsibilities and Rights Journal* 26(3):195–216.

Desmarais, Serge, and James Curtis. 2001. "Gender and Perceived Income Entitlement among Full-Time Workers: Analyses for Canadian National Samples, 1984 and 1994." *Basic and Applied Social Psychology* 23(3):157–68.

Deutsch, Morton. 1985. *Distributive Justice: A Social-Psychological Perspective*. New Haven, CT: Yale University Press.

Eagly, Alice H., and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review* 109(3):573–98.

England, Paula. 1992. *Comparable Worth: Theories and Evidence*. New York: Aldine de Gruyter.

European Commission. 2015. "Company Income Reports, Austria." (Website managed by the European Commission's Directorate-General for Justice and Consumers.) Luxembourg: European Commission. Retrieved February 21, 2016 (http://ec.europa.eu/justice/gender-equality/gender-pay-gap/depth-info/tools-gpg/reports-austria/index_en.htm).

Eurostat. 2015. "Gender Pay Gap in Unadjusted Form." (Main Tables on Labour Market Earnings.) Luxembourg: European Commission. Retrieved November 11, 2016 (http://ec.europa.eu/eurostat/web/products-datasets/product?code=tsdsc340).

Faia, Michael A. 1980. "The Vagaries of the Vignette World: A Comment on Alves and Rossi." *American Journal of Sociology* 85(4):951–54.

Fang, Hanming, and Andrea Moro. 2011. "Theories of Statistical Discrimination and Affirmative Action: A Survey." Pp. 133–200 in *Handbook of Social Economics*, Vol. 1A, edited by J. Benhabib, A. Bisin, and M. O. Jackson. Amsterdam: North-Holland, Elsevier.

Fehr, Ernst, Lorenz Goette, and Christian Zehnder. 2009. "A Behavioral Account of the Labor Market: The Role of Fairness Concerns." *Annual Review of Economics* 1:355–84.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7(2):117–40.

Fişek, M. Hamit, and Stuart J. Hysom. 2008. "Status Characteristics and Reward Expectations: A Test of a Theory of Justice in Two Cultures." *Social Science Research* 37(3):769–86.

Fishbein, Martin, and Icek Ajzen. 2010. *Predicting and Changing Behavior: The Reasoned Action Approach*. New York: Psychology Press.

Foschi, Martha. 1996. "Double Standards in the Evaluation of Men and Women." *Social Psychology Quarterly* 59(3):237–54.

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

Foschi, Martha, Larissa Lai, and Kirsten Sigerson. 1994. "Gender and Double Standards in the Assessment of Job Applicants." *Social Psychology Quarterly* 57(4):326–39.

Gangl, Markus, and Andrea Ziefle. 2009. "Motherhood, Labor Force Behavior, and Women's Careers: An Empirical Assessment of the Wage Penalty for Motherhood in Britain, Germany, and the United States." *Demography* 46(2):341–69.

Ganzeboom, Harry B. G., and Donald J. Treiman. 1996. "Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations." *Social Science Research* 25(3):201–39.

Gartner, Hermann, and Thomas Hinz. 2009. "Geschlechtsspezifische Lohnungleichheit in Betrieben, Berufen und Jobzellen (1993–2006)." *Berliner Journal für Soziologie* 19(4):557–75.

Gatskova, Kseniia. 2013. "Distributive Justice Attitudes in Ukraine: Need, Desert or Social Minimum?" *Communist and Post-Communist Studies* 46(2):227–41.

Gibson, Donald E., and Barbara S. Lawrence. 2010. "Women's and Men's Career Referents: How Gender Composition and Comparison Level Shape Career Expectations." *Organization Science* 21(6):1159–75.

Hedström, Peter, and Richard Swedberg. 1998. "Social Mechanisms: An Introductory Essay." Pp. 1–31 in *Social Mechanisms. An Analytical Approach to Social Theory*, edited by P. Hedström and R. Swedberg. Cambridge, UK: Cambridge University Press.

Homans, George C. 1973. "Fundamental Social Processes." Pp. 549–94 in *Sociology: An Introduction*, 2nd ed. Edited by N. J. Smelser. New York: John Wiley and Sons.

Hox, Joop J., Ita G. G. Kreft, and Piet L. J. Hermkens. 1991. "The Analysis of Factorial Surveys." *Sociological Methods and Research* 19(4):493–510.

Husted, Bryan W., and Robert Folger. 2004. "Fairness and Transaction Costs: The Contribution of Organizational Justice Theory to an Integrative Model of Economic Organization." *Organization Science* 15(6):719–29.

Jann, Ben. 2005. *Erwerbsarbeit, Einkommen und Geschlecht. Studien zum Schweizer Arbeitsmarkt*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Jasso, Guillermina. 1980. "A New Theory of Distributive Justice." *American Sociological Review* 45(1):3–32.

Jasso, Guillermina. 1990. "Methods for the Theoretical and Empirical Analysis of Comparison Processes." *Sociological Methodology* 20:369–419.

Jasso, Guillermina. 1996. "Exploring the Reciprocal Relations between Theoretical and Empirical Work: The Case of the Justice Evaluation Function (Paper in Honor of Robert K. Merton)." *Sociological Methods and Research* 24(3):253–303.

Jasso, Guillermina. 2006. "Factorial Survey Methods for Studying Beliefs and Judgments." *Sociological Methods and Research* 34(3):334–423.

Jasso, Guillermina. 2012. "Safeguarding Justice Research." *Sociological Methods and Research* 41(1):217–39.

Jasso, Guillermina, and Peter H. Rossi. 1977. "Distributive Justice and Earned Income." *American Sociological Review* 42(4):639–51.

Jasso, Guillermina, and Murray Webster Jr. 1997. "Double Standards in Just Earnings for Male and Female Workers." *Social Psychology Quarterly* 60(1):66–78.

Jasso, Guillermina, and Murray Webster Jr. 1999. "Assessing the Gender Gap in Just Earnings and Its Underlying Mechanisms." *Social Psychology Quarterly* 62(4):367–80.

Kalkhoff, Will, and Shane R. Thye. 2006. "Expectation States Theory and Research: New Observations from Meta-Analysis." *Sociological Methods and Research* 35(2):219–49.

Kluegel, James R., and Eliot R. Smith. 1986. *Beliefs About Inequality: Americans' Views of What Is and What Ought to Be*. New York: Aldine de Gruyter.

Krysan, Maria, Mick P. Couper, Reynolds Farley, and Tyrone A. Forman. 2009. "Does Race Matter in Neighborhood Preferences? Results from a Video Experiment." *American Journal of Sociology* 115(2):527–59.

Major, Brenda. 1989. "Gender Differences in Comparisons and Entitlement: Implications for Comparable Worth." *Journal of Social Issues* 45(4):99–115.

Major, Brenda. 1994. "From Social Inequality to Personal Entitlement: The Role of Social Comparisons, Legitimacy Appraisals, and Group Membership." Pp. 293–355 in *Advances in Experimental Social Psychology*, Vol. 26, edited by M. P. Zanna. San Diego, CA: Academic Press.

Major, Brenda, and Blythe Forcey. 1985. "Social Comparisons and Pay Evaluations: Preferences for Same-Sex and Same-Job Wage Comparisons." *Journal of Experimental Social Psychology* 21(4):393–405.

Major, Brenda, and Ellen Konar. 1984. "An Investigation of Sex Differences in Pay Expectations and Their Possible Causes." *Academy of Management Journal* 27(4):777–92.

Major, Brenda, Dean B. McFarlin, and Diana Gagnon. 1984. "Overworked and Underpaid: On the Nature of Gender Differences in Personal Entitlement." *Journal of Personality and Social Psychology* 47(6):1399–1412.

Major, Brenda, and Maria Testa. 1989. "Social Comparison Processes and Judgments of Entitlement and Satisfaction." *Journal of Experimental Social Psychology* 25(2):101–20.

Mandel, Hadas, and Moshe Semyonov. 2005. "Family Policies, Wage Structures, and Gender Gaps: Sources of Earnings Inequality in 20 Countries." *American Sociological Review* 70(6):949–67.

Markovsky, Barry, and Kimmo Eriksson. 2012. "Comparing Direct and Indirect Measures of Just Rewards." *Sociological Methods and Research* 41(1):199–216.

Martell, Richard F., David M. Lane, and Cynthia Emrich. 1996. "Male-Female Differences: A Computer Simulation." *American Psychologist* 51(2):157–58.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Melamed, David. 2012. "Deriving Equity from Expectations: A Cross-Cultural Evaluation." *Social Science Research* 41(1):170–81.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*, Vol. 2. New York: National Bureau of Economic Research.

Moore, Dahlia. 1991. "Entitlement and Justice Evaluations: Who Should Get More, and Why." *Social Psychology Quarterly* 54(3):208–23.

Mueller, Charles W., and Sang-Wook Kim. 2008. "The Contented Female Worker: Still a Paradox?" *Advances in Group Processes* 25:117–49.

Mutz, Diana C. 2011. *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press.

Neumark, David. 2012. "Detecting Discrimination in Audit and Correspondence Studies." *Journal of Human Resources* 47(4):1128–57.

Nielsen, Helena Skyt, Marianne Simonsen, and Mette Verner. 2004. "Does the Gap in Family-Friendly Policies Drive the Family Gap?" *Scandinavian Journal of Economics* 106(4):721–44.

Pager, Devah, and Lincoln Quillian. 2005. "Walking the Talk? What Employers Say Versus What They Do." *American Sociological Review* 70(3):355–80.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62(4):659–61.

Quillian, Lincoln. 2006. "New Approaches to Understanding Racial Prejudice and Discrimination." *Annual Review of Sociology* 32:299–328.

Rashotte, Lisa Slattery, and Murray Webster Jr. 2005. "Gender Status Beliefs." *Social Science Research* 34(3):618–33.

Ridgeway, Cecilia L. 1997. "Interaction and the Conservation of Gender Inequality: Considering Employment." *American Sociological Review* 62(2):218–35.

Ridgeway, Cecilia L. 2001. "Gender, Status, and Leadership." *Journal of Social Issues* 57(4):637–55.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford, UK: Oxford University Press.

Ridgeway, Cecilia L., and Shelley J. Correll. 2004. "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations." *Gender and Society* 18(4):510–31.

Sauer, Carsten, Katrin Auspurg, Thomas Hinz, and Stefan Liebig. 2011. "The Application of Factorial Surveys in General Population Samples: The Effects of Respondent Age and Education on Response Times and Response Consistency." *Survey Research Methods* 5(3):89–102.

Sauer, Carsten, Katrin Auspurg, Thomas Hinz, Stefan Liebig, and Jürgen Schupp. 2014. "Method Effects in Factorial Surveys: An Analysis of Respondents' Comments, Interviewers' Assessments, and Response Behavior." *SOEPpapers on Multidisciplinary Panel Data Research* # 629. Berlin: German Institute for Economic Research (DIW).

Schaeffer, Nora Cate. 1991. "Hardly Ever or Constantly? Group Comparisons Using Vague Quantifiers." *Public Opinion Quarterly* 55(3):395–423.

Schauer, Frederick F. 2003. *Profiles, Probabilities, and Stereotypes*. Cambridge, MA: Belknap Press of Harvard University Press.

Shamon, Hawal, and Hermann Dülmer. 2014. "Raising the Question on 'Who Should Get What?' Again: On the Importance of Ideal and Existential Standards." *Social Justice Research* 27(3):340–68.

Shepelak, Norma J., and Duane F. Alwin. 1986. "Beliefs about Inequality and Perceptions of Distributive Justice." *American Sociological Review* 51(1):30–46.

Smolny, Werner, and Matthias Kirbach. 2011. "Wage Differentials between East and West Germany: Are They Related to the Location or to the People?" *Applied Economics Letters* 18(9):873–79.

Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69(5):797–811.

Thébaud, Sarah. 2015. "Status Beliefs and the Spirit of Capitalism: Accounting for Gender Biases in Entrepreneurship and Innovation." *Social Forces* 94(1):61–86.

Tomaskovic-Devey, Donald, and Sheryl Skaggs. 1999. "An Establishment-Level Test of the Statistical Discrimination Hypothesis." *Work and Occupations* 26(4):422–45.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study

(SOEP) – Scope, Evolution and Enhancements." *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 127(1):139–69.

Walster, Elaine, Ellen Berscheid, and G. William Walster. 1973. "New Directions in Equity Research." *Journal of Personality and Social Psychology* 25(2):151–76.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.

Katrin Auspurg holds a full professorship in sociology (specializing in quantitative empirical research) at the Department of Sociology at the LMU Munich, Germany. She is co-author of *Factorial Survey Experiments* (Sage Publications, Series Quantitative Applications in Social Sciences). She investigates how inequalities in the labor market and the family intersect and reinforce each other. In addition, her current projects advance innovative experimental or survey methods that allow the testing of mechanisms that cause social inequalities or subtle forms of discrimination.

Thomas Hinz holds a full professorship in empirical social research and survey methodology at the Department of Sociology at the University of Konstanz, Germany. His research interests include social inequalities and discrimination in markets. He has investigated the development of the gender wage gap in Germany in cooperation with the Institute of Employment Research. He is co-author of *Factorial Survey Experiments* (Sage Publications, Series Quantitative Applications in Social Sciences).

Carsten Sauer is a Radboud Excellence Initiative Fellow (postdoc) at the Department of Sociology at Radboud University Nijmegen, Netherlands. His research interests include the explanation of behavior, social inequality and justice, and quantitative research methods (especially factorial surveys).