The Gender of Opportunity: How Gendered Job Titles Affect Job Seeker Attraction.*

Paul M. Gorny, Petra Nieken, and Martin Trenkle

March 31, 2026

Abstract

Language carries information. In labor markets, even minimal linguistic cues in job ads can function as signals that update job seekers' beliefs about organizational culture and thereby shape application behavior. We study this mechanism using three complementary datasets. In a large-scale field experiment on a major German job platform, switching job titles from the generic masculine to a gender-fair form increases female applications in *Business & Management* by roughly 50%, with no decline in male applications. This effect is mirrored by a significant increase in female clicks, indicating that the signal operates already at the stage of initial consideration. Effects are absent in *Marketing & Sales* and in *IT & Development* overall, but emerge for IT positions posted by non-IT companies–consistent with a belief-updating mechanism that requires sufficiently elastic priors. An online study with hiring experts shows that gender-fair title usage is driven more by genuine support for gender equality than by strategic motives, suggesting that the titles are credible employer-type signals. A laboratory eye-tracking experiment with job seekers confirms that behavioral responses reflect belief-based evaluation rather than automatic visual salience. Together, the evidence demonstrates that language operates as a market signal in recruitment: a virtually costless three-character change can meaningfully expand the applicant pool when prior beliefs have room for updating.

Keywords: Gender, Labor, Recruitment, Bias, Discrimination, Gender-Fair Language

JEL: C93, J16, J20, J62, J71, Z13

*Gorny, Nieken: Institute of Management, Karlsruhe Institute of Technology, Nieken: CESifo Research Network Fellow, Trenkle: Workwise GmbH, Karlsruhe. This project was funded as part of the Excellence Strategy of the German Federal and State Governments. This project was approved by the ethics committee of the Karlsruhe Institute of Technology. Our RCT was preregistered at the AEA RCT Registry (Gorny and Nieken, 2023) and our online and lab studies were jointly preregistered on OSF (#6sepv). We thank Iris Kesternich, Lydia Mechtenberg, Gerd Mülheusser, and Niklas Wallmeyer, as well as the participants of the Workshop on Gender in Adaptive Design 2023, the Microeconomics Research Seminar at the University of Hamburg, the OB Research Seminar at HEC Lausanne, the ECON Brown Bag Seminar at the Karlsruhe Institute of Technology, the HeiKaMaX 2023 at the University of Mannheim, and the Economics Lunch Seminar at the University of Regensburg for helpful comments, and Alexa Becker, Walter Gottfried, Sergiu Panainte, Marius Spanka, Fabian Wuest, and Rebecca Zimmer for excellent research assistance. All remaining errors are our own.

1 Introduction

Companies routinely struggle to fill positions with qualified candidates, yet many suitable applicants never apply. This is not simply a supply problem; qualified workers exist, but something in the recruitment process prevents them from entering the pipeline. One important margin is the application stage itself: even among observably similar candidates, men and women do not apply to the same jobs, and women often apply more narrowly, especially when opportunities are more ambiguous, more demanding, or associated with male-dominated job categories (Cortés et al., 2023; Coffman et al., 2024; Fluchtmann et al., 2024). Labor market matching depends critically on whether job seekers correctly interpret employer signals and self-select into consideration. When those signals are noisy or systematically misread, talent is misallocated not because it is absent, but because it never enters the applicant pool (Hsieh et al., 2019). We study whether a single linguistic choice in the job title, a three-character difference between the generic masculine and its gender-fair equivalent in German, can shift beliefs about organizational culture and thereby broaden the applicant pool.

Both the prior literature and simple economic reasoning invite skepticism about large effects. Job seekers are making consequential decisions, choosing where to invest their time, reputation, and career steps, and they plausibly care far more about salary, job content, and career prospects than about three characters in a title. If anything, one might expect that sophisticated labor market participants see through surface-level linguistic signals, particularly in a language like German where gender-fair forms are culturally visible and actively debated, making them easy to dismiss as symbolic rather than informative. Studies examining gendered wording in job advertisements find mixed and often modest effects on application behavior (Gaucher et al., 2011; Abraham et al., 2024), and the most closely related field experiment, conducted on a Spanish-language tech job platform, finds an average treatment effect of zero (Del Carpio and Fujiwara, 2026). A further concern cuts in the opposite direction: even if gender-fair titles attract more female applicants, they might simultaneously deter male applicants, leaving firms with a reshuffled rather than expanded pool. The anticipation of a male backlash is sufficiently widespread to plausibly discourage employers from adopting gender-fair language even when they would otherwise be inclined to do so. Taken together, the existing evidence offers little reason to expect that a minimal linguistic intervention in a job title would move application behavior in any consistent or economically meaningful way.

This paper, however, provides evidence from three complementary datasets suggesting why substantial effects may nonetheless arise: a randomized field experiment on a German online job platform, an online study of hiring experts, and a controlled laboratory study with job seekers and eye-tracking. The following theoretical considerations help organize that evidence.[1] Job seekers enter a given job category with priors about the kind of organizational culture they are likely to encounter. A gender-fair title changes behavior only if it works as a credible and informative signal. If priors are too pessimistic, as in strongly male-dominated job categories, the signal may not be credible enough to overturn them (Pan, 2015). If priors are already favorable, the signal may not be informative enough to change behavior. The largest effects should therefore arise in a sweetspot region of priors in which gender-fair language is both credible and informative, so that the title can move a vacancy across the threshold for consideration.

Our field-experimental results line up closely with this theory. In the field experiment, we assigned job ads on a German online job platform to either the generic masculine or its gender-fair equivalent in the title. Our data span three job categories that differ markedly in their gender composition: *IT & Development* (ID), *Business & Management* (BM), and *Marketing & Sales* (MS). In BM, which appears closest to the sweetspot region, gender-fair titles increase total applications and female applications in particular, with female applications increasing by roughly 50%. The effect is mirrored by a significant increase in female clicks, indicating that the treatment already changes which vacancies enter job seekers' consideration sets. In ID, the average effect is close to zero, yet a revealing pattern emerges: IT jobs posted by non-IT companies attract significantly more female applicants when the gender-fair title is used, consistent with the idea that priors are less extreme in this subgroup. In MS, where the female labor share is largest, we find no comparable effects. Pooling across all job categories yields a null average effect, which is itself informative because it highlights why the category design matters: the signal works where priors leave room for updating, not as a uniform average treatment across all recruitment contexts. Crucially, we find no evidence of a male backlash in any job category: male application rates are unaffected by the treatment.

A signaling argument also requires that the signal be credible. Thus, job seekers must have reason to believe that employers using gender-fair titles are more likely to be inclusive rather than merely mimicking inclusiveness. We test this directly with an online study of hiring experts, professionals actively involved in recruitment decisions, who evaluated job-ad design choices and made incen-

[1] We provide a simple formalization of this intuition in Appendix A.

tivized title choices. Two findings emerge. First, hiring experts expect gender-fair titles to appeal to female applicants, but they also widely anticipate a male backlash that does not materialize in our field data. Second, and more importantly for the credibility of the signal, the decision to use a gender-fair title in an incentivized job-ad design task is driven not only by strategic beliefs about its effectiveness but even more strongly by hiring experts' underlying support for gender equality. This is precisely the condition under which a signal is informative: employers who send it are more likely to possess the characteristic it conveys. The gap between expected and actual backlash also helps explain why this low-cost intervention can remain underused.

The laboratory study speaks to the other side of the mechanism, namely whether the title is processed as informative rather than merely salient. To identify this, we complement the field experiment with a controlled laboratory study that combines an incentivized job-search task with screen-based eye-tracking. This allows us to observe not only what job seekers decide but how they process information on the way to that decision, distinguishing between attention allocation and evaluative judgment. The key pattern from the lab is a dissociation between attention and choice. In *Marketing & Sales*, gender-fair titles can increase fixation shares without translating into more clicks. In *Business & Management*, and partly at the dyadic level in *IT & Development*, female clicks increase without corresponding increases in fixation shares. If the effect operated mainly through visual salience, we would expect attention and choice to move more closely together. In our data, however, they do not systematically do so. This pattern is more consistent with a belief-updating process: job seekers are not drawn to gender-fair titles merely because they are visually salient, but because, once encountered, the titles shift beliefs about the workplace environment in a way that crosses the threshold for consideration. This interpretation is further supported by the finding that male job seekers in *IT & Development* click less on gender-fair titles in the lab, yet show no reduction in their intent to apply.

Each of the three studies in this paper is individually informative, but their value is cumulative. The field experiment shows where the effect appears, the hiring-expert study shows why the signal can be credible, and the laboratory study shows that the response is more consistent with belief updating than with mere salience. Together, the three studies trace the full chain from the employer's language choice, through the job seeker's cognitive response, to actual application behavior.

Our findings contribute to three strands of literature. We add to the economics of recruitment and labor-market matching by providing evidence that linguistic signals in job titles can gener-

ate economically meaningful shifts in applicant-pool composition (Barron and Bishop, 1985; Banfi and Villena-Roldan, 2019). We contribute to the literature on gendered language and economic behavior by identifying the mechanism, belief updating at the stage of initial consideration rather than low-level attention capture, through which language affects consequential decisions, and by establishing the conditions under which such effects emerge (Jakiela and Ozier, 2020; Del Carpio and Fujiwara, 2026). Finally, we speak to the literature on gender differences in job search by showing that language in job titles can relax a margin at which qualified women apply too narrowly, thereby improving access to talent for firms (Cortés et al., 2023; Coffman et al., 2024; Fluchtmann et al., 2024).

Closest to our paper is Del Carpio and Fujiwara (2026), who study gender-neutral language in job ads on a Spanish-language tech platform. Their zero average treatment effect and strong spillover results are highly complementary to our evidence: in both settings, language matters when it is informative relative to the surrounding informational environment. Our intervention is milder and more scalable because we manipulate only the title, we cover three job categories rather than one, and we trace the mechanism with both hiring experts and eye-tracking data.

Our findings have direct practical implications. For firms facing skill shortages, gender-fair job titles are a rare intervention that is virtually costless, generates no measurable male backlash, and in our setting expands the female talent pool without evidence of lower application quality. More broadly, the results show that part of the gender gap in hiring arises not only from who is qualified, but from who enters the application funnel in the first place. The remainder of this paper is structured as follows. In Section 2, we describe and discuss our evidence from the RCT. Section 3 presents the hiring expert study and the laboratory experiment. We discuss findings, implications, and limitations in Section 4. Section 5 concludes.

2 Company Data from a Randomized Controlled Trial

2.1 Study Environment and Procedures

We will start by describing the institutional setting and experimental manipulation. We first introduce the job platform and ad structure. Then, we outline the randomized experimental design and data collection.

2.1.1 The Job Platform

Our randomized field experiment was conducted in cooperation with Workwise GmbH, a large German online job platform providing HR and recruiting services to companies across the coun-

try.[2] A key feature of this platform is that companies pay only if a vacancy is successfully filled. All job seekers use the platform free of charge but must register to apply. Thus, all applications are processed and recorded within the platform's database. This institutional setting allows us to observe the full set of applications submitted through the platform. In addition to hosting job ads, Workwise actively distributes (spreads) ads to external partner websites to increase outreach. These spreading ads account for roughly 60 percent of total traffic and constitute the environment in which our experimental manipulation takes place.

Workwise classifies each job ad into ten broad job categories. Job ads follow a standardized structure: a header containing the job title, the company name, an application button, followed by a textual job description, information on benefits, and a structured list of required skills. Because our experimental manipulation is confined exclusively to the job title, this standardized layout ensures that any observed variation in job seeker behavior can be unambiguously attributed to our manipulation rather than to formatting differences.

2.1.2 Experimental Design

To study whether gender-fair job titles affect job seeker behavior, we implemented a randomized controlled trial (RCT). The study was preregistered at the AEA RCT registry (Gorny and Nieken, 2023). Our study encompasses two treatments. In the *Baseline*, job ads were displayed with generic masculine job titles. In the *Gender-Fair* treatment, jobs were advertised using a gender-fair title (e.g., the generic masculine "Entwickler" was changed to "Entwickler:in").

We used the colon (":") to construct gender-fair job titles, as it is compatible with screen readers and thus accessible to visually impaired users, while being increasingly adopted in German institutional contexts.[3]

Importantly, the treatment was implemented via job ad *spreading*, i.e., posts made on other platforms to increase the reach of the job ads. These spreading ads account for roughly 60% of the job platform's traffic.

We focus on three job categories that differ substantially in the female employment share: "IT & Development (ID)," "Business & Management (BM)," and "Marketing & Sales (MS)."[4] These categories allow us to examine if the treatment effects vary with the baseline gender composition.

[2] The company was founded under the name Campusjäger.

[3] Alternative forms of gender-fair language in German include the gender star ("*"), the underscore ("_"), double forms (e.g., "Entwickler und Entwicklerinnen"), and neutral formulations (e.g., "Entwickelnde").

[4] According to Institut für Arbeitsmarkt- und Berufsforschung (IAB) (2025), the current female employment shares are 16.3% in ID, 43.9% in BM, and 54.0% in MS.

Our primary outcome is the number of applications per job ad, including the number of applications from females.

2.1.3 Data Collection

We defined eligible jobs as (i) full-time jobs that are (ii) not already advertised using gender-fair titles by the hiring company and that (iii) fall within the categories "IT & Development," "Business & Management," or "Marketing & Sales." After eligibility screening by the account managers, the job ads entered the RCT and were randomly assigned to one of the two treatments with equal probability.

We started data collection in September 2021. We preregistered to stop data collection once each of the six treatment and job category (2×3) contained at least 300 observations. Data collection ended on November 15, 2022. yielding a raw sample of 2,051 job ads.

2.2 Data Preparation, Estimation Strategy, and Summary Statistics

Next, we describe the analytical sample, key variables, and empirical strategy.

2.2.1 Sample Selection and Unit of Observation

Each job ad assigned to one of the job-category-treatment combinations counted toward our target of 300 observations per cell. Table 1 reports the distribution of observations across treatments and job categories.

	ID	**BM**	**MS**	Total
Baseline	353	367	284	1,004
Gender-Fair	295	334	265	894
Total	648	701	549	1,898

Table 1: Observations by treatment and job category

In some cases, job ads were assigned to a treatment but never posted online (e.g., deleted or not activated). We exclude these ads from the analysis. After these exclusions, the analytical sample consists of 1,898 observations. The unit of observation is the job ad. For each ad, we track all job seeker behavior during the period in which it was active (or until the end of our sampling window, whichever occurred first). Balance tests (reported in Tables A.3–A.5 in the Online Appendix) show no systematic differences between the Baseline and the Gender-Fair ads in observable characteristics. The only statistically significant difference is a slightly higher number of required skills in the Gender-Fair treatment in BM. Given that more required skills are negatively correlated with application rates, especially among females, this imbalance would bias the estimated treatment

effects downward.

2.2.2 Variables of Interest

Our treatment variable *Gender-Fair* equals one if the job ad title was displayed in a gender-fair form and zero otherwise. We focus on three job categories. The dummy *ID* indicates whether or not a job was in the category *IT & Development*. We will use the symbol 💻 for this category in the figures. Similarly, the dummies *BM* and *MS* indicate whether the job underlying the job ad was in the category *Business & Management* or *Marketing & Sales*, respectively. We will use the symbols 💼 and 📣 for these categories in the figures, respectively.

Our primary variables measure job seeker behavior at the job-ad level. *No. of applications* give the total number of completed applications, *No. of female applications*, the number of female applications and the female applicant ratio is given by dividing female applications by total applications and denoted by *Share of female applications*. Due to the required platform registration, the job seeker's gender is reliably observed for completed applications.

To study the mechanisms along the application funnel, we additionally analyze the count of clicks that led job seekers to the job ad, given by the variable *Number of clicks*. The count of clicks by registered female users on the platform is stored in *No. of registered female clicks*. Further controls derived from text mining and LLM-based classification of job ad content are described in Appendix C.2.

Job-ad characteristics vary across job categories (Table 16 in Appendix D), including the number of words, remote-work options, and baseline use of gender-fair wording. These differences reflect structurally distinct job environments. We therefore estimate category-specific treatment effects, as preregistered.

2.3 Hypotheses

Proponents of gender-fair language argue that the generic masculine can implicitly signal a less inclusive organizational culture. This could discourage individuals whose gender identity is not reflected in the linguistic form. Gender-fair titles in job ads, by contrast, can serve as a signal for inclusivity.

Persistent occupational segregation means that women in male-dominated fields often work in environments where their perspectives are underrepresented and their needs less well accommodated (Goldin, 2014; Blau and Kahn, 2017). An inclusive organizational culture, therefore, carries asymmetric value: women, who bear the costs of underrepresentation, benefit disproportion-

ately from employers that credibly signal inclusivity, whereas men–lacking comparable exposure to such costs–are less responsive to these signals or may even perceive them with ambivalence. Gender-fair job titles can function as precisely such a signal, updating job seekers' beliefs about an employer's organizational culture. The decision to click on the job ad and apply then not only depends on the listed skills and offers in the job ad, but also on updated beliefs about the workplace environment. A minimal formalization of this signaling mechanism, in the spirit of Coffman et al. (2024), is provided in Appendix A.

Job ads communicate more than tasks and requirements; they convey information about organizational culture (Rafaeli, 2000; Ganesan et al., 2018). Using GF language in the title may therefore act as a credible signal of inclusivity. Job seekers who value inclusivity, particularly women, should respond more strongly to this signal. This leads to the following preregistered hypotheses.

Hypothesis RCT.1. Job ads using gender-fair titles receive more female applications.

We also examine whether the nominal increase in female applications translates into a higher female applicant ratio.

Hypothesis RCT.2. Job ads using gender-fair titles exhibit a higher female applicant ratio.

Our preregistered stratification into three job categories–each with distinct baseline levels of female clicks and applications–reflects our concern that these underlying differences may interact with potential treatment effects. We therefore account for these category-specific differences in our empirical specification, while keeping the direction and magnitude of any resulting heterogeneity in treatment effects as an exploratory matter.

Because our three job categories differ substantially in baseline gender representation and job-ad characteristics–and hence in the prior beliefs and signal credibility that govern the model's predictions–we allow treatment effects to vary by category, consistent with our preregistration.

2.4 Results

As discussed above, the three job categories differ significantly along several pre-treatment dimensions. Consistent with our theoretical framework, we estimate category-specific effects. Regression results are displayed in Tables 2 and 3.[5]

In Specification (1), which pools all categories, the estimated treatment effect is small and statistically insignificant. In Specification (2) to (5), we allow the treatment effect to vary by job category,

[5] A detailed description of our empirical strategy is provided in Appendix B.

Dep. Var.: Number of applications	(1)	(2)	(3)	(4)	(5)
Gender-Fair	-0.350	2.638	3.135*	2.923*	3.118*
	(1.159)	(1.845)	(1.776)	(1.762)	(1.800)
ID		5.040**	2.939	0.470	0.840
		(2.007)	(1.922)	(2.256)	(2.236)
MS		6.552***	4.124**	2.449	2.350
		(2.040)	(2.054)	(2.188)	(2.194)
Gender-Fair×ID		-3.072	-3.336	-3.831	-4.243
		(2.746)	(2.613)	(2.569)	(2.617)
Gender-Fair×MS		-6.567**	-5.922**	-5.092*	-5.305**
		(2.845)	(2.700)	(2.675)	(2.688)
Constant	12.419***	8.789***	4.408	7.198	2.651
	(0.773)	(1.194)	(4.968)	(7.409)	(7.262)
p(ID)	.	0.830	0.917	0.632	0.556
p(MS)	.	0.070	0.175	0.287	0.280
Company	✗	✗	✓	✓	✓
Application	✗	✗	✗	✓	✓
Job ad	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.001	0.008	0.011	0.013
Observations	1898	1898	1898	1898	1898

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID $= 0$ and Gender-Fair + Gender-Fair × MS $= 0$ respectively. The complete tables with all covariate coefficients can be found in Tables A.8a through A.8e in the Online Appendix.

Table 2: Tobit regression on the number of applications received per job ad

with BM as the reference group. In the fully controlled Specification (5), gender-fair titles significantly increase the number of applications in BM. By contrast, the treatment effect for ID and MS is small and statistically not different from zero. This pattern is consistent with non-monotonic treatment effects across job categories.[6]

Table 3 reports the number of female applications. As in the previous table, pooling across job categories yields no statistically significant treatment effect. Once we allow for job category-specific effects, however, a clear pattern emerges.

In the fully controlled Specification (5), gender-fair titles increase female applications in BM by approximately 2.5 additional applications. Given that the average number of female applications in the Baseline was 4.902, this corresponds to an increase of nearly 50%.

There is no evidence of treatment effects in ID and MS.

[6]To obtain the treatment effects in the categories ID and MS and test if they are significantly different from zero, we need to consider the linear sum of Gender-Fair + Gender-Fair × ID and Gender-Fair + Gender-Fair × MS, respectively. The p-values associated with the test of whether they are equal to zero are reported in the lower panel of Table 2.

Dep. Var.: Number of female applications	(1)	(2)	(3)	(4)	(5)
Gender-Fair	0.365	2.136*	2.490**	2.322**	2.389**
	(0.661)	(1.239)	(1.182)	(1.172)	(1.199)
ID		-3.206***	-3.873***	-4.366***	-4.126***
		(1.080)	(1.128)	(1.277)	(1.296)
MS		3.406***	2.167*	1.354	1.284
		(1.157)	(1.151)	(1.273)	(1.282)
Gender-Fair×ID		-1.841	-2.212	-2.311	-2.518*
		(1.518)	(1.464)	(1.440)	(1.460)
Gender-Fair×MS		-4.244**	-4.047**	-3.204**	-3.302**
		(1.712)	(1.635)	(1.618)	(1.628)
Constant	1.933***	2.073***	0.604	5.394	2.706
	(0.419)	(0.712)	(3.741)	(4.835)	(4.744)
p(ID)	.	0.738	0.750	0.989	0.881
p(MS)	.	0.074	0.170	0.430	0.412
Company	✗	✗	✓	✓	✓
Application	✗	✗	✗	✓	✓
Job ad	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.005	0.017	0.023	0.025
Observations	1898	1898	1898	1898	1898

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.9a through A.9e in the Online Appendix.

Table 3: Tobit regression on the number of female applications received per job ad

Result RCT.1. Gender-fair job titles increase the number of applications in BM. This is mainly driven by a substantial increase in the number of female applications ($\approx 50\%$). The treatment does not affect the number of applications in ID and MS.

Pooling across job categories (see Specification (1) in the previous tables) shows no significant treatment effect on the number of applications or the number of female applications.
This is similar when considering the share of female applications when pooling over all job categories.[7]

Result RCT.2. Gender-fair job titles have no statistically significant impact on the number of applications, female application numbers, or the female applicant ratio when pooling across categories.

2.5 Further Analysis and Mechanisms

The previous section concluded our preregistered analyses. We now examine additional evidence speaking to the signaling mechanism underlying our results. In our model, gender-fair titles act

[7] Ads without any application are excluded from this estimation. Our results can be found in Section A.1.3 of the Online Appendix.

as signals of a more inclusive workplace culture and therefore affect job seekers' beliefs before the application decision is made. This mechanism, together with our experimental design, implies two key empirical patterns: first, belief updating should be strongest when the signal is credible, and priors are intermediate; second, due to our treatment focusing on titles, effects should emerge already at the attention stage, reflected in clicking behavior. In addition, to imply the observed absence of a male backlash across job categories, our model requires that male job seekers are largely insensitive to the signal ($g_M \approx 0$). The following analyses examine these considerations more thoroughly.

2.5.1 Signal Credibility and Tipping Points in the IT Industry

Previous studies documented 'tipping points' in heavily male-dominated fields such as STEM or IT (Pan, 2015; Del Carpio and Fujiwara, 2026).

Dep. Var.: Number of applications	(1)	(2)	(3)	(4)	(5)
Gender-Fair	-0.421	2.985	3.059	2.369	2.774
	(2.019)	(2.326)	(2.123)	(2.246)	(2.276)
IT company		8.519***	10.603	12.014	11.493
		(3.142)	(7.216)	(7.410)	(7.269)
Gender-Fair×IT company		-6.498	-5.075	-5.075	-6.521
		(4.073)	(3.749)	(3.934)	(4.256)
Constant	13.755***	9.341***	-2.181	-7.339	-12.163
	(1.408)	(1.601)	(9.286)	(15.056)	(14.778)
p(ID Company)	.	0.284	0.495	0.364	0.245
Company	✗	✗	✓	✓	✓
Application	✗	✗	✗	✓	✓
Job ad	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.002	0.014	0.017	0.019
Observations	648	648	648	648	648

Robust standard errors in parentheses; $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$
Note: The p-values in the lower panel, p(IT Company), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × IT Company = 0. The complete tables with all covariate coefficients can be found in Tables A.14a through A.14d in the Online Appendix.

Table 4: Tobit regression on the number of applications a job ad in ID received depending on whether the advertising company was in an IT company (ID job ads only)

In environments where the representation of females is extremely low and this fact is salient to all job seekers, only considerably strong interventions might change the gender balance away from that extreme. Thus, gender-fair job titles might carry limited informational content and thus induce little belief updating.

Within our ID jobs, there is variation in terms of the companies' industry affiliation. Roughly half

of the IT & Development *jobs* are advertised by IT *companies*, whereas the other half are advertised by non-IT companies (51.08 % IT, 48.92 % non-IT).[8]

Table 4 shows the results for the number of applications considering only ID jobs and allowing for heterogeneous effects between jobs advertised by IT and non-IT companies. Whereas we see that IT companies, on average, received significantly more applications, in line with our main findings, there is no indication of differences in our treatment effect.

Dep. Var.: Number of female applications	(1)	(2)	(3)	(4)	(5)
Gender-Fair	0.125	1.812*	1.782*	1.703*	1.856*
	(0.719)	(1.009)	(0.956)	(0.967)	(0.972)
IT company		3.560***	4.278**	3.974*	3.822*
		(1.070)	(2.065)	(2.264)	(2.139)
Gender-Fair×IT company		-3.143**	-2.526*	-2.519*	-3.053**
		(1.473)	(1.365)	(1.378)	(1.463)
Constant	0.579	-1.290*	-2.052	-0.467	-3.250
	(0.448)	(0.713)	(4.042)	(6.008)	(6.008)
p(ID Company)	.	0.198	0.432	0.381	0.220
Company	✗	✗	✓	✓	✓
Application	✗	✗	✗	✓	✓
Job ad	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.004	0.022	0.033	0.036
Observations	648	648	648	648	648

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(IT company), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × IT company $= 0$. The complete tables with all covariate coefficients can be found in Tables A.15a through A.15d in the Online Appendix.

Table 5: Tobit regression on the number of female applications a job ad in ID received, depending on whether the advertising company was an IT company (ID job ads only)

Turning to the number of female applications reported in Table 5, we observe a positive treatment effect ($p = 0.057$) for non-IT companies, but not for IT companies. This increase amounts to approximately 1.856 additional female applications, an effect size of nearly 58.9 % relative to the Baseline within the ID job category. This pattern aligns with our signaling interpretation: where extreme male dominance is most salient (IT companies), the gender-fair title appears less effective. When the prior belief is less extreme (e.g., non-IT companies hiring IT jobs), the same signal induces measurable belief updating.

[8] Additional information on the share of IT companies among job advertisers can also be found in Table A.1 in the Online Appendix.

2.5.2 Consideration Decisions and the Extensive Margin

We next examine whether the treatment affects the probability that the job ad receives at least one application (*Non-deserted* job ads). Across treatments, 90.89 % of the job ads receive at least one application. There is a nominal but statistically insignificant ($p = 0.474$) increase of 0.95 percentage points in the Gender-Fair treatment, with an average of 91.39 %, compared to 90.44 % in the Baseline. The result is similar when considering female applications only. The mean share of non-deserted ads across both treatments is 70.71 %. There is a nominal but statistically not significant ($p = 0.272$) increase of 2.30 percentage points in the Gender-Fair treatment, with an average of 71.92 %, compared to 69.62 % in the Baseline.[9]

2.5.3 Attention and Belief Updating

The job ad titles were not only visible on the job ad page itself but also on the spreading platforms when receiving a list from a keyword search or in email alerts.

Dep. Var.: Number of clicks	(1)	(2)	(3)	(4)	(5)
Gender-Fair	-4.055	11.658	15.414	15.897*	17.446*
	(6.199)	(10.217)	(9.913)	(9.609)	(9.787)
ID		23.566**	10.495	1.187	4.679
		(9.978)	(9.501)	(10.915)	(10.815)
MS		40.792***	30.050***	27.558**	28.074**
		(11.660)	(11.375)	(11.562)	(11.499)
Gender-Fair×ID		-17.910	-21.043	-24.243*	-25.795*
		(14.156)	(13.536)	(13.459)	(13.562)
Gender-Fair×MS		-33.159**	-29.724**	-28.084*	-29.360**
		(16.005)	(15.084)	(14.647)	(14.695)
Constant	77.536***	57.704***	21.587	31.875	0.381
	(4.446)	(6.266)	(21.598)	(32.617)	(32.553)
p(ID)	.	0.523	0.546	0.369	0.371
p(MS)	.	0.081	0.210	0.277	0.283
Company	✗	✗	✓	✓	✓
Application	✗	✗	✗	✓	✓
Job ad	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.001	0.007	0.009	0.010
Observations	1898	1898	1898	1898	1898

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.12a through A.12e in the Online Appendix.

Table 6: Tobit regression on the number of clicks a job ad received

If gender-fair titles function as signals, belief updating should occur at the first exposure, affecting

[9] Regression results are reported in Tables A.10a through A.11e in the Online Appendix.

clicking behavior before the application decision is made.

Table 6 reports the effects of total clicks: In the fully controlled specification, we see a positive effect in job category BM but no effect for the other two categories, mirroring the previous findings.

Table 7 reports the number of female clicks. Gender-fair job titles significantly increased the number of female clicks on the job ad in BM, whereas there was no significant effect in the other two categories.

Dep. Var.: Number of female clicks	(1)	(2)	(3)	(4)	(5)
Gender-Fair	-0.543	8.301	10.326**	10.032**	10.853**
	(2.882)	(5.169)	(4.999)	(4.852)	(4.965)
ID		-5.818	-10.866***	-13.126***	-11.347**
		(3.874)	(4.067)	(4.521)	(4.543)
MS		21.771***	16.703***	14.683**	14.967***
		(5.884)	(5.601)	(5.810)	(5.802)
Gender-Fair×ID		-10.007*	-11.813**	-12.260**	-13.072**
		(6.002)	(5.790)	(5.706)	(5.812)
Gender-Fair×MS		-20.110**	-18.917**	-16.734**	-17.438**
		(8.189)	(7.665)	(7.437)	(7.465)
Constant	29.606***	25.483***	17.447	27.274	11.199
	(1.904)	(3.085)	(11.763)	(16.992)	(17.060)
p(ID)	.	0.575	0.625	0.473	0.473
p(MS)	.	0.063	0.139	0.237	0.240
Company	✗	✗	✓	✓	✓
Application	✗	✗	✗	✓	✓
Job ad	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.002	0.011	0.014	0.015
Observations	1898	1898	1898	1898	1898

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.13a through A.13e in the Online Appendix.

Table 7: Tobit regression on the number of female clicks a job ad received

Comparing effect sizes, we see that the overall increase of about 17 clicks in BM was mainly driven by the roughly 11 additional female clicks. This is a sizable effect translating to roughly 40 % more female clicks in the Gender-Fair treatment compared to the Baseline, all other things equal.

The parallel pattern in clicks and the number of applications supports a signaling mechanism: gender-fair titles shift female applicants' beliefs at the initial attention stage, and these updated beliefs translate into higher numbers of applications.

2.5.4 Application Quality

A potential concern is that increases in applications may come at the expense of application quality. Our data allow us to examine this directly using the platform-generated application score, a machine-learning-based measure of applicant fit that incorporates both the number of job requirements met and the submitted application materials. Estimating OLS regressions using the same specifications as in our main analysis, we find no statistically significant effects of the treatment on application quality. The corresponding results are reported in Tables A.19a through A.19e in the Online Appendix.

3 Hiring Experts' and Job Seekers' Behavior and Beliefs

In the following, we will refer to participants in our hiring expert study as *hiring experts* and to participants in our lab experiment as *job seekers*.

3.1 Hiring Expert Study

To understand how individuals with hiring experience anticipate the effects of gender-fair language, and whether these beliefs align with our field results, we conducted an online study with hiring experts. The study serves two purposes (i) to elicit the expert's belief about how gender-fair job titles affect applications from different groups, and (ii) to test if their beliefs are consistent with a signaling-based interpretation. We recruited German, Austrian, and Swiss participants via Prolific who reported fluency in German and prior experience in hiring decisions.[10]

3.1.1 Study Design

At the beginning of the study, participants confirmed their hiring experience to validate the prescreening. Those without hiring experience were excluded. Participants then selected the job category most closely matching their background (IT & Development, Business & Management, or Marketing & Sales) and classified their hiring experience using the job taxonomy of the German Federal Employment Agency (KldB; Bundesagentur für Arbeit, 2021). This allows us to align the hiring experts beliefs with the categories used in the field experiment.

Belief Elicitation. In Part 1, hiring experts stated their beliefs about how six job-ad design features affect the number of applications: (i) remote versus on-site work, (ii) flexible versus fixed hours, (iii) team-oriented versus competitive environment, (iv) gender-fair versus generic-masculine

[10] Eligibility was determined using Prolific's prescreen question: "Do you have any experience in making hiring decisions (i.e., have you been responsible for hiring job candidates)?"

job titles, (v) gender-fair versus generic-masculine language in the job description, and (vi) informal ("Du") versus formal ("Sie") address.[11] Beliefs were elicited separately for eight applicant groups, including female and male applicants.[12]

Incentivized Design Choices. In Part 2, hiring experts made actual job ad design choices. Each hiring expert was randomly assigned to one standardized job ad from the category corresponding to their experience. The job ads were harmonized versions of the RCT ads, sampled from different parts of the application distribution to ensure variation while maintaining comparability. The hiring experts chose among five binary dimensions, including gender-fair versus generic masculine titles.[13] To incentivize choices, the hiring experts were informed that one configuration could be implemented in a subsequent job seeker lab study. Their potential bonus depended on whether at least one job seeker applied to their ad. The probability of receiving the bonus depended on the job seekers application decision combined with a probabilistic component.[14] This incentive structure encourages the hiring experts to reveal their true beliefs about what attracts job seekers, rather than merely expressing attitudes.

Task Environment and Attitudes In Part 3, hiring experts selected two parameters of the subsequent real-effort task environment: whether the task instructions would be framed in formal or informal language, and whether performance feedback would be presented in a competitive or team-oriented manner. Finally, in Part 4, the hiring experts completed a short survey on attitudes toward language change and the representation of non-binary individuals, followed by demographic questions.

3.1.2 Procedures

In total, 178 hiring experts participated. The study was programmed in oTree (Chen et al., 2016). The median completion time was approximately 21-22 minutes. All participants received a fixed

[11] We use a symmetric pairwise framing that reduces experimenter-demand effects (Zizzo, 2010; Mummolo and Peterson, 2019) and limits respondent fatigue by reducing the number of evaluations.

[12] The groups were overall applicants; younger (<30) and older (>50) applicants; applicants for whom German is not the first language and applicants from abroad; and male, female, as well as applicants who do not (fully) identify with either gender category.

[13] The five dimensions were work arrangement remote/flexible vs. on-site/time-structured), work-environment framing (team-oriented vs. competitive), job-title language (gender-fair vs. generic-masculine), main-text language (gender-fair vs. generic-masculine), and formality of address (informal "Du" vs. formal "Sie").

[14] Hiring experts received an excerpt of the job-seeker instructions and were informed about the real-effort task to understand the context and consequences of their design decisions. The job seeker's performance in the task did not affect the expert's bonus.

payment of £3.00, corresponding to an hourly rate of about £8.50. Including potential bonus payments, the average total payment amounted to £3.47.

3.1.3 Results

Table 8 reports summary statistics for hiring experts across the three job categories.

Variable	ID	BM	MS	Total	
Age	36.74 (7.86)	37.44 (11.17)	37.16 (12.01)	37.17 (10.65)	p=0.849
Female	0.20 (0.40)	0.32 (0.47)	0.40 (0.49)	0.31 (0.47)	p=0.087
Migration background	0.35 (0.48)	0.32 (0.47)	0.25 (0.44)	0.31 (0.46)	p=0.557
Encountered equality topics at work	0.72 (0.46)	0.60 (0.49)	0.53 (0.50)	0.61 (0.49)	p=0.148

Note: The figures show means with standard deviations in parentheses. The p-values in the last column represent the results from Kruskal-Wallis (KW) and Fisher exact tests (F).

Table 8: Summary statistics across job categories

The gender distribution mirrors the expected pattern based on the underlying share of women in each occupation as well as the share of female applications for each job category in our RCT, supporting the external validity of the hiring expert sample: 19.57 % female in IT & Development (ID), 32.47 % in Business & Management (BM), and 40.00 % in Marketing & Sales (MS) with marginally statistically significant differences across job categories ($p = 0.087$, Kruskal-Wallis test).[15]

We find no significant differences in overall job ad configurations across job categories ($p = 0.827$, joint F-test after a multivariate regression of all design choices on job-category dummies). Thus, systematic category differences in design choices do not drive the patterns reported below.

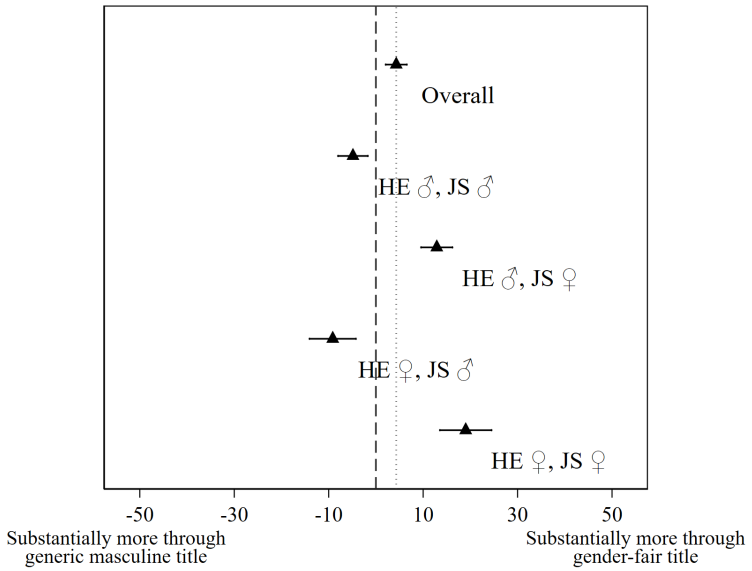
Overall, hiring experts expected GF titles attract more applications ($+0.161$sd, $p = 0.033$, t).[16] However, this aggregated belief masks a strongly gender-differentiated expectation. When elicited separately by job seeker gender, hiring experts expect gender fair titles to increase female applications and decrease male applications, implying a positive anticipated gender gap in application behavior. On average, hiring experts expected the difference between female and male applicants to be 21.028 percentage points, which translates into 0.848 standard deviations ($p < 0.001$).[17] In other words, hiring experts widely believed in a male backlash combined with a female boost.

[15] Distributions of additional controls can be found in Tables A.21 through A.29 in the Online Appendix.

[16] The p-values refer to t-tests of whether the mean is equal to zero, whereas the beliefs are provided in standard deviations (sd) of the overall mean from zero due to the fact that we elicited hiring experts' beliefs on a subjective scale.

[17] Figure 2 in the appendix summarizes these results.

This pattern varies somewhat by job category: the positive overall beliefs are primarily driven by BM ($+0.361$sd, $p = 0.002$, t) rather than MS ($+0.095$sd, $p = 0.485$, t), and ID. In ID, hiring experts stated a negative (though insignificant) expectation (-0.040sd, $p = 0.789$, t). Figure 1 further shows that both female and male hiring experts share this directional belief. However, female experts predict significantly stronger effects in both directions, with larger positive effects for female and more pronounced negative effects for male job seekers. These differences are statistically significant ($p = 0.009$) and highlight that while both genders share the same directional beliefs, female hiring experts hold them more strongly.



Note: Triangles show mean responses on how gender-fair titles affect male and female application numbers; whiskers denote 95% confidence intervals. The dashed and dotted lines indicate zero and the overall mean, respectively.

Figure 1: Hiring experts' beliefs about how different titles affect the number of applications by Female hiring expert (HE) gender and job seeker (JS) gender

We next study whether these beliefs translate into design choices. Table 9 reports Probit regressions regarding the decision to use a gender-fair job title. Specifications (1) and (2) show that hiring experts who believe that GF titles increase the number of applications (both male and female) are significantly more likely to choose them.[18] The marginal effect of one standard deviation increase about the belief about female applications is 3.150 percentage points and 10.184 for male applications. At first sight, this pattern points to a strategic signalling motive: Hiring experts use gender-fair titles when they believe that these improve the appeal of job ads to female applicants.

[18]Naturally, results are the same when using beliefs about total applications instead.

Dep. Var.: GF Title Design	(1)	(2)	(3)	(4)
Effect Belief of gender fair title (female)	0.016***	0.016***	0.007	0.007
	(0.005)	(0.006)	(0.006)	(0.006)
Effect Belief of gender fair title (male)	0.012**	0.019***	0.009	0.014**
	(0.006)	(0.007)	(0.006)	(0.007)
Female		0.319		0.166
		(0.262)		(0.271)
Support for gender equality			0.403***	0.432***
			(0.092)	(0.092)
Constant	0.294**	0.790	-1.558***	-1.410
	(0.132)	(1.037)	(0.427)	(1.117)
Controls	✗	✓	✗	✓
Pseudo R^2	0.071	0.195	0.173	0.289
Observations	178	178	178	178

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Controls are job-category dummies, age, education, mother tongue, income, migration background, responses to the time-flexibility and remote-work vignettes, number of recruits previously hired, company language, encountered equality topics, company-size dummies, whether the company is based in Germany, and individual risk preferences. The complete tables with all covariate coefficients can be found in Tables A.31a through A.31b in the Online Appendix.

Table 9: The signalling value of GF titles (Probit regressions)

However, once we include support for gender equality (with higher values indicating greater support for gender equality) in Specifications (3) and (4), the picture changes. The coefficients on the belief about female applications decline in magnitude and lose statistical significance in Specification (4). The marginal effect of a one standard deviation increase in the belief about female applications reduces to 3.294 percentage points. In contrast, the positive coefficient on the gender-equality measure is large, highly significant, and stable across specifications.[19] A one standard deviation increase in support for gender equality is associated with a 14.577 percentage points higher propensity to choose the gender-fair title.

Beliefs about male applications, by contrast, remain relatively stable when attitudes are included. The marginal effect remains at 9.203 percentage points. Overall, normative orientation toward gender equality appears to be the dominant predictor of GF-title adoption.

Descriptive differences in design choices across job categories are reported in Appendix Table 17. Overall, the results imply that the decision to use GF titles reflects both a strategic and a normative motivation, but that the latter dominates. Hiring experts with stronger support for gender-diversity are substantially more likely to use GF titles. Thus, the signalling value of GF titles appears to be "real:" They are credible indicators of an underlying preference for diversity rather than mere rhetorical devices. Because their use stems from genuine attitudes, job seekers can

[19] This result is qualitatively robust to running OLS models with the same set of specifications.

plausibly interpret GF titles as authentic signals of such attitudes on the hiring side.

3.2 Job Seeker Study

In the ensuing lab experiment, participants from a student participant pool were invited to act as potential *job seekers*.

3.2.1 Study Design

Prior to the on-site study, eligible participants[20] completed a short online prescreening survey. After providing informed consent, they reported their degree program and field of study, indicated the job category they would most likely apply to after graduation, and assessed how well this category aligned with their career plans. This procedure mirrors the categorization in the field experiment, ensuring that the job seekers are exposed to job ads that are personally relevant. This allows us to induce job-category-specific expectations and signal credibility in line with our theoretical framework. We used this information to balance invitations to the on-site sessions across the three job categories corresponding to our experimental conditions. Participants also stated their preferred job category, absent constraints.

We further administered the UAS work preferences module (Mas and Pallais, 2017), which elicits reservation wage premia for flexible scheduling and workplace flexibility using two standardized vignette choices. Finally, participants reported a range of standard demographics.[21] Gender information was used to ensure balanced participation and enable heterogeneity analyses.

The main job seeker study was conducted on-site and comprised three parts. The procedure was identical across conditions, except that participants were assigned to one of three categories (IT & Development, Business & Management, or Marketing & Sales) based on their stated preference. This category-specific assignment directly parallels the field experiment, where the treatment effects were heterogeneous across job categories. In each condition, participants viewed twelve job ads, nine of which matched their preferred category. To assess attentiveness, we included three additional ads: one from each of the two other categories and one ad entirely unrelated to any job seeker's preferences. The order was randomized, and the study was programmed using oTree (Chen et al., 2016).

[20] Eligible participants were fluent in German, had unrestricted or corrected vision, and studied in one of the following fields: Industrial Engineering, Computer Science, Mechanical Engineering, Electrical Engineering and Information Technology, Business Informatics, Mechatronics and Information Technology, Mathematics, Science, Media and Communication, Business Mathematics, Digital Economics, Geodesy and Geoinformatics, Technomathematics, Technology and Management in Construction Operations, Mobility and Infrastructure, Computer Science Teaching, and Remote Sensing and Geoinformatics.

[21] These were gender, age, mother tongue, additional languages, and migration background.

After receiving introductory information and providing informed consent, participants were informed that there are three parts to the study.

Job Seeking Task. The task contained two pages. On Page 1, participants could toggle between four screens, each containing three job *titles* only. Participants were required to select at least four and at most twelve titles. Crucially, this stage isolates the pure title-level signal before any additional job information is revealed. The titles differed in the use of gender-fair language. For each participant, we randomly sampled twelve hiring experts from the hiring-expert study. Nine were drawn from the participant's preferred job category and used to generate the corresponding nine relevant ads. One hiring expert was drawn from each of the two other categories and one from the full pool to generate the three non-relevant ads. This procedure tightly links this study to the hiring expert study; each title reflects an actual design choice made by a real hiring expert. For the nine relevant ads, we imposed that between two and seven titles used gender-fair language. The first four selections were costless; additional selections entailed convex costs. The cost structure induces meaningful trade-offs and prevents indiscriminate clicking, thereby revealing preference-based job-seeker behavior.

After selecting at least four ads, participants proceeded to Page 2, where they could inspect the full advertisements (intro, tasks, profile, offer) and rank their four most preferred ads (you can think of this as "shortlisting" ads). The final application was assigned probabilistically: 40% (rank 1), 30% (rank 2), 20% (rank 3), and 10% (rank 4).

Participants were informed that the ads were based on design choices made by hiring experts in a previous study and that the environment for the subsequent real-effort task in Part 2 would depend on the matched hiring expert design. Thus, job titles potentially not only signal stylistic language but underlying employer type represented by the hiring expert, similar to the field experiment.

Importantly, the participants did not observe the specific work-environment design before ranking.

Real Effort Task. In Part 2, participants completed a real-effort task whose framing depended on the matched expert's design choices: instructions used either formal or informal language, and feedback was framed either competitively (revealing relative rank) or team-oriented (contribution to a team of three). The other team members had participated in an earlier session and did not ob-

serve the participant's performance. This implements the consequence of choosing an employer: a job-title signal is linked to an actual real-effort task environment, preserving the signal's incentive compatibility. The task required alternating key presses (A and B) (DellaVigna and Pope, 2018), with each correct pair yielding one point. Participants earned a €3 bonus if they achieved at least 200 points within 90 seconds.

Survey and Eye-Tracking. Finally, in Part 3, participants completed a survey eliciting social preferences and attitudes toward gendered language. We used eye-tracking throughout Part 1 to record gaze patterns. This allows us to observe visual attention allocation when the signal is received. Participants received a fixed payment of €10, minus selection costs, an additional €2 for passing attention checks, and a potential €3 performance bonus.

3.3 Procedures

We invited participants via ORSEE (Greiner, 2015). Of 203 presurvey participants, 136 qualified for our lab study.[22] The total average payment for participants who took part in the lab study amounted to approximately €17.65, and the average lab session lasted 35 minutes.[23] Eye movements were recorded using Tobii 4C eye-trackers, which were calibrated individually. Only participants with unrestricted or appropriately corrected vision were eligible to participate.[24]
Importantly, this study was conducted after completion of the field experiment, and the participants had no knowledge of the results.

3.4 Hypotheses

The theoretical mechanism underlying our RCT is that gender-fair job titles serve as a signal about the workplace environment. The hiring-expert study showed that such titles are more likely to be chosen by experts with a stronger gender equality attitude and are perceived as gender-differentiated signals in hiring. The lab study, therefore, investigates the job-seeker side of the signaling equilibrium: do job seekers attend to and act upon this signal?
We deliberately structured the job seeker lab study to parallel key features of the field context while maintaining a controlled setting.
If gender fair job titles function as a signal about an inclusive organizational culture, they must

[22] Summary statistics can be found in Table 18 in Appendix D.3.1.

[23] The average payment includes the presurvey payment. The average for the 203 participants was approximately €12.81.

[24] Tobii 4C devices can be used with glasses, provided that the lenses are not excessively thick or reflective. For one participant, the calibration failed due to excessive deviation in gaze alignment (Strabismus), and this participant was therefore excluded from the study before the start of the experimental tasks.

first capture attention. In a competitive visual environment, attention allocation reflects perceived relevance (Loewenstein and Wojtowicz, 2025). This should be reflected in the relative gaze share directed at the title region.

Hypothesis Job Seeker.1. Job ads using gender-fair job titles receive more visual attention than ads using the generic masculine (measured by the relative share of fixations on the title).

Second, if gender-fair titles reduce psychological distance and signal a more inclusive organizational culture, women in particular should be more inclined to select such ads during the search stage of the application task.

Hypothesis Job Seeker.2. Women are more likely to select job ads with gender-fair titles than those with generic masculine titles on the search page of the application task.

These hypotheses focus on early-stage behavioral responses, attention, and selection because they represent the first points at which job seekers react to linguistic signals in the hiring process, allowing us to compare these mechanisms directly with the field data.

3.4.1 Results

In our main analysis, we restructure the data from the job-seeker level to the job-ad level such that each observation represents one unique job-ad configuration. This aggregation mirrors the ad-level structure of the field experiment, allowing for a direct comparison of treatment effects across studies. Specifically, we generate all possible combinations of the five binary design features (gender-fair title, gender-fair text, informal address, team orientation, and flexible work) for each of the 36 job ads and retain only those configurations that were actually displayed to job seekers. We then merge information on displayed ads, clicks, and rankings from the search page, aggregating job-seeker characteristics to the ad level using weighted means by the number of displays, clicks, and ranks. This produces, for each ad configuration, measures of exposure, attractiveness, and ranking outcomes, together with the average characteristics of the job seekers who viewed, clicked, or ranked it. The resulting dataset thus contains one observation per displayed job-ad configuration, enriched with the corresponding design features, aggregated behavioral outcomes, and job-seeker covariates, forming the basis for the ad-level analyses paralleling our analysis of the RCT data. The treatment effects occur at the ad level.

Table 10 reports the effects of gender-fair titles on the number of job-ad clicks across all job seekers. While point estimates for BM are positive and those for ID and MS are close to zero, none reach statistical significance. Importantly, the cross-category pattern broadly mirrors the field evidence.

Dep. Var.: Number of clicks	(1)	(2)	(3)	(4)	(5)	(6)
Gender-Fair	-0.176	-0.016	-0.027	-0.015	-0.018	-0.009
	(0.250)	(0.467)	(0.296)	(0.296)	(0.298)	(0.301)
ID		0.431	0.089	0.102	0.095	0.103
		(0.421)	(0.253)	(0.255)	(0.261)	(0.265)
MS		0.031	0.205	0.189	0.185	0.268
		(0.384)	(0.256)	(0.257)	(0.259)	(0.262)
Other		-3.243***	-1.402***	-1.405***	-1.406***	-1.332***
		(0.663)	(0.408)	(0.409)	(0.412)	(0.413)
Gender-Fair×ID		0.007	0.061	0.029	0.034	0.021
		(0.636)	(0.384)	(0.386)	(0.391)	(0.392)
Gender-Fair×MS		-0.265	-0.254	-0.257	-0.255	-0.297
		(0.624)	(0.392)	(0.390)	(0.392)	(0.401)
Gender-Fair×Other		-0.282	-0.440	-0.451	-0.454	-0.602
		(0.898)	(0.564)	(0.560)	(0.560)	(0.551)
Constant	0.142	0.408	-0.862***	-1.696***	-1.728**	-1.622
	(0.227)	(0.309)	(0.303)	(0.577)	(0.770)	(1.379)
p(GF: ID)	.	0.984	0.899	0.958	0.953	0.966
p(GF: MS)	.	0.489	0.297	0.310	0.311	0.264
Exposure	✗	✗	✓	✓	✓	✓
Demographics	✗	✗	✗	✓	✓	✓
Attitudes	✗	✗	✗	✗	✓	✓
Behavioral	✗	✗	✗	✗	✗	✓
Pseudo R^2	0.001	0.050	0.231	0.233	0.233	0.235
Observations	517	517	517	517	517	517

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.32a through A.32b in the Online Appendix.

Table 10: Tobit regressions on the number of clicks for the job ads displayed in the job seeker study

Positive estimates in BM and attenuated effect in ID and MS. Results are robust across controls for exposure, demographics, attitudes, and behavioral measures. The consistently negative and significant coefficient for the "Other" category indicates that job seekers correctly identified irrelevant job ads and adjusted their clicking behavior. This reassures that job seekers processed the information meaningfully rather than clicking randomly.

Table 11 presents clicks by female job seekers. The positive and statistically significant coefficient for the gender-fair title in *Business & Management* (the omitted statistical baseline) shows that ads with gender-fair titles received approximately half an additional female click, on average, relative to ads without gender-fair titles (0.550 clicks in the full specification). This effect is robust across all specifications. In contrast, the interaction terms for *IT & Development* and *Marketing & Sales* are not statistically significant. As before, the negative coefficient for "Other" jobs confirms that job seekers paid attention. This is in line with the field experiment results: gender-fair titles increase female engagement in BM but not in the other two categories. Thus, the job seeker study repro-

duces the category-specific female behavior observed in the field, strengthening the interpretation that gender-fair titles are relevant signals rather than generic attention shifters.

Dep. Var.: Number of female clicks	(1)	(2)	(3)	(4)	(5)	(6)
Gender-Fair	0.265	0.701*	0.550*	0.582*	0.535	0.550*
	(0.237)	(0.404)	(0.326)	(0.322)	(0.326)	(0.325)
ID		0.296	0.109	0.259	0.118	0.060
		(0.410)	(0.327)	(0.335)	(0.346)	(0.340)
MS		0.392	0.460	0.470	0.451	0.695**
		(0.398)	(0.331)	(0.332)	(0.334)	(0.332)
Other		-1.866***	-0.867	-0.869	-0.986*	-0.831
		(0.687)	(0.554)	(0.530)	(0.531)	(0.531)
Gender-Fair×ID		-0.168	-0.064	-0.107	-0.030	-0.038
		(0.567)	(0.438)	(0.436)	(0.437)	(0.429)
Gender-Fair×MS		-0.972	-0.894*	-0.972**	-1.036**	-1.177**
		(0.599)	(0.476)	(0.477)	(0.476)	(0.489)
Gender-Fair×Other		-0.561	-0.660	-0.629	-0.590	-0.932
		(0.874)	(0.706)	(0.688)	(0.685)	(0.666)
Constant	-0.187	-0.165	-0.859**	-0.597	-2.453**	-2.014
	(0.188)	(0.303)	(0.407)	(0.874)	(1.154)	(1.807)
p(GF: ID)	.	0.183	0.139	0.151	0.119	0.100
p(GF: MS)	.	0.540	0.356	0.302	0.184	0.108
Exposure	✗	✗	✓	✓	✓	✓
Demographics	✗	✗	✗	✓	✓	✓
Attitudes	✗	✗	✗	✗	✓	✓
Behavioral	✗	✗	✗	✗	✗	✓
Pseudo R^2	0.001	0.039	0.136	0.141	0.146	0.161
Observations	360	360	360	360	360	360

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.33a through A.33b in the Online Appendix.

Table 11: Tobit regressions on the number of clicks by women for the job ads displayed in the job seeker study

Result Job Seeker.1. A deviation from the generic masculine in job titles increases the number of female clicks in BM. This effect does not translate to an increased number of total clicks and is absent in ID and MS.

Table 12 reports the effects of gender-fair titles on the relative share of fixations directed at job titles on the search page. Across specifications, the estimates are directionally in line with the results for the number of clicks but statistically not significant. Thus, gender-fair job titles do not systematically alter early-stage visual attention at the aggregate level.

Table 13 presents fixations by female job seekers. Across specifications, the estimated effects generally move in the same direction as the click results for *Business & Management* and *IT & De-*

Dep. Var.: Share of fixations	(1)	(2)	(3)	(4)	(5)	(6)
Gender-Fair	2.612	2.025	0.916	0.871	0.824	0.794
	(2.767)	(5.733)	(1.784)	(1.784)	(1.794)	(1.814)
ID		3.858	-1.506	-1.676	-1.655	-1.405
		(4.992)	(2.168)	(2.161)	(2.193)	(2.210)
MS		-3.624	-0.070	-0.054	-0.145	-0.169
		(4.067)	(1.705)	(1.709)	(1.715)	(1.721)
Other		-23.496***	-4.840***	-4.882***	-4.795***	-4.679***
		(3.521)	(1.481)	(1.480)	(1.485)	(1.485)
Gender-Fair×ID		0.341	2.179	2.192	2.286	2.217
		(8.055)	(2.870)	(2.907)	(2.942)	(2.978)
Gender-Fair×MS		3.242	3.236	3.331	3.482	3.638
		(7.516)	(2.549)	(2.548)	(2.578)	(2.590)
Gender-Fair×Other		1.172	-1.518	-1.558	-1.606	-1.711
		(6.014)	(2.120)	(2.131)	(2.129)	(2.161)
Constant	22.514***	26.127***	-1.630	-2.866	-1.168	-6.025
	(2.058)	(3.467)	(1.803)	(2.931)	(3.517)	(5.955)
p(GF: ID)	.	0.675	0.187	0.196	0.191	0.211
p(GF: MS)	.	0.280	0.026	0.025	0.022	0.018
Exposure	✗	✗	✓	✓	✓	✓
Demographics	✗	✗	✗	✓	✓	✓
Attitudes	✗	✗	✗	✗	✓	✓
Behavioral	✗	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.009	0.221	0.221	0.221	0.221
Observations	517	517	517	517	517	517

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.34a through A.34b in the Online Appendix.

Table 12: Tobit regressions on the share of fixations on job titles displayed in the job seeker study

velopment, whereas no such alignment is visible for *Marketing & Sales*. However, none of these coefficients is statistically significant. Hence, at the job ad level, we find no evidence that female job seekers allocate more visual attention to gender-fair titles.

Taken together, these results yield an important insight. Gender-fair titles increased female clicking behavior in BM, replicating the field evidence, but do not measurably shift fixation shares in this job category. This suggests that the signaling effect operates at the level of evaluation and selection rather than through automatic, low-level attention capture, which, in the case of MS, does not affect behavior.

This dissociation between gaze patterns and behavior is consistent with a belief updating mechanism rather than a purely salience-driven attention mechanism. If gender-fair titles were merely perceived as visual oddities, we would expect significant fixation differences and behavioral differences for the same group of job seekers.

Result Job Seeker.2. Gender-fair titles do not increase the relative share of fixations directed at

Dep. Var.: Share of female fixations	(1)	(2)	(3)	(4)	(5)	(6)
Gender-Fair	2.479	3.517	1.952	2.003	1.755	1.210
	(2.030)	(4.095)	(2.186)	(2.184)	(2.139)	(2.128)
ID		2.290	-0.053	0.164	-0.696	-1.283
		(3.265)	(2.239)	(2.255)	(2.271)	(2.302)
MS		-0.194	1.073	1.422	1.590	2.758
		(2.917)	(2.087)	(2.104)	(2.015)	(2.076)
Other		-13.982***	-2.919*	-2.577	-3.460*	-2.736
		(2.468)	(1.706)	(1.769)	(1.783)	(1.823)
Gender-Fair×ID		-1.469	-0.692	-0.562	-0.083	0.556
		(5.434)	(3.141)	(3.165)	(3.109)	(3.085)
Gender-Fair×MS		1.800	1.634	1.396	0.728	0.682
		(5.693)	(3.231)	(3.201)	(3.164)	(3.184)
Gender-Fair×Other		-2.382	-3.709	-3.833	-3.546	-4.297*
		(4.466)	(2.521)	(2.540)	(2.518)	(2.539)
Constant	16.555***	17.624***	0.986	5.749	-10.948**	-16.296*
	(1.188)	(2.248)	(1.975)	(3.751)	(4.907)	(8.323)
p(GF: ID)	.	0.567	0.593	0.544	0.474	0.440
p(GF: MS)	.	0.179	0.132	0.153	0.297	0.434
Exposure	✗	✗	✓	✓	✓	✓
Demographics	✗	✗	✗	✓	✓	✓
Attitudes	✗	✗	✗	✗	✓	✓
Behavioral	✗	✗	✗	✗	✗	✓
Pseudo R^2	0.000	0.011	0.133	0.133	0.136	0.141
Observations	360	360	360	360	360	360

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.35a through A.35b in the Online Appendix.

Table 13: Tobit regressions on the share of female fixations on job titles displayed in the job seeker study

job titles, neither overall nor for female job seekers.

3.4.2 Further Analyses

Building on the ad-level results, we now turn to analyses at the individual level. To this end, we restructure the data into job seeker-job ad dyads, yielding a dataset of $1,632$ observations (136 job seekers $\times$ 12 job ads). This dyadic setup allows us to jointly account for variation in individual and job-ad characteristics within a unified analysis. Since we previously analyzed female clicks and applications, we run these regressions split for male and female job seekers to avoid three-way interactions between the gender-fair titles, job categories, and the job seeker's gender.[25]

Table 14 reports Probit estimates for whether a female job seeker clicked on a job ad. The Gender-Fair coefficient in *Business & Management* is positive and marginally statistically significant across

[25]This is mainly for representation purposes; the results are robust to a specification with such interactions.

Dep. Var.: Clicked	(1)	(2)	(3)	(4)	(5)
Gender-Fair	0.125*	0.220*	0.260*	0.279*	0.275*
	(0.075)	(0.128)	(0.142)	(0.145)	(0.149)
ID		-0.043	-0.041	-0.044	-0.073
		(0.123)	(0.124)	(0.127)	(0.135)
MS		0.123	0.134	0.133	0.142
		(0.144)	(0.145)	(0.144)	(0.159)
Other		-0.741**	-0.744**	-0.734**	-0.743**
		(0.364)	(0.364)	(0.364)	(0.371)
Gender-Fair×ID		0.097	0.090	0.095	0.119
		(0.196)	(0.203)	(0.204)	(0.208)
Gender-Fair×MS		-0.330	-0.336	-0.343	-0.358
		(0.231)	(0.235)	(0.235)	(0.241)
Gender-Fair×Other		-0.263	-0.260	-0.278	-0.282
		(0.435)	(0.436)	(0.441)	(0.445)
Constant	-0.431***	-0.410***	-0.359***	-0.268*	0.459
	(0.048)	(0.083)	(0.122)	(0.140)	(0.348)
p(GF: ID)	.	0.037	0.044	0.037	0.032
p(GF: MS)	.	0.515	0.669	0.722	0.644
p(GF: Other)	.	0.917	1.000	0.997	0.986
Ad text	✗	✗	✓	✓	✓
Exposure	✗	✗	✗	✓	✓
Individual controls	✗	✗	✗	✗	✓
Pseudo R^2	0.002	0.027	0.027	0.028	0.030
Observations	804	804	804	804	804

Cluster-robust SEs (individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.40a through A.40b in the Online Appendix.

Table 14: Probit regressions on whether a job ad was clicked by a female in the job seeker study within a dyad

all five specifications ($p = 0.064$ in the most saturated model), confirming the main result from Table 11 at the individual level. Extending that finding, the joint test p (GF: ID) is statistically significant ($p = 0.032$), indicating that the positive effect on female clicking generalizes to *IT & Development* as well. No significant effect is found in *Marketing & Sales*. In line with our main results, these estimates underscore that gender-fair language increases female engagement with job ads, and suggest the effect extends beyond BM.

Table 15 examines whether a female job seeker placed a job ad in her shortlist (measuring the intent to apply).

The pattern closely mirrors the click results: the Gender-Fair coefficient in *Business & Management* is positive and marginally statistically significant ($p = 0.052$), confirming that gender-fair titles

Dep. Var.: Ranked	(1)	(2)	(3)	(4)	(5)
Gender-Fair	0.110	0.227*	0.273*	0.288**	0.291*
	(0.075)	(0.127)	(0.142)	(0.146)	(0.150)
ID		-0.088	-0.086	-0.087	-0.099
		(0.120)	(0.125)	(0.128)	(0.141)
MS		0.109	0.121	0.120	0.124
		(0.138)	(0.140)	(0.140)	(0.153)
Other		-0.698*	-0.702*	-0.694*	-0.692*
		(0.361)	(0.361)	(0.362)	(0.366)
Gender-Fair×ID		0.040	0.024	0.028	0.031
		(0.194)	(0.202)	(0.202)	(0.208)
Gender-Fair×MS		-0.331	-0.339	-0.344	-0.355
		(0.229)	(0.232)	(0.233)	(0.238)
Gender-Fair×Other		-0.399	-0.397	-0.411	-0.427
		(0.441)	(0.441)	(0.444)	(0.448)
Constant	-0.490***	-0.452***	-0.407***	-0.336***	-0.200
	(0.041)	(0.079)	(0.119)	(0.119)	(0.167)
p(GF: ID)	.	0.065	0.088	0.080	0.078
p(GF: MS)	.	0.548	0.715	0.756	0.726
p(GF: Other)	.	0.686	0.777	0.777	0.758
Ad text	✗	✗	✓	✓	✓
Exposure	✗	✗	✗	✓	✓
Individual controls	✗	✗	✗	✗	✓
Pseudo R^2	0.001	0.026	0.027	0.027	0.028
Observations	804	804	804	804	804

Cluster-robust SEs (individual) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Note: The p-values in the lower panel, p(GF: ID) and p(GF: MS), correspond to two-sided tests of the linear hypotheses Gender-Fair + Gender-Fair × ID = 0 and Gender-Fair + Gender-Fair × MS = 0 respectively. The complete tables with all covariate coefficients can be found in Tables A.42a through A.42b in the Online Appendix.

Table 15: Probit regressions on whether a job ad was ranked by a female in the job seeker study within a dyad

also increase females' intent to apply to BM ads. The joint test $p(GF : ID)$ is marginally statistically significant ($p = 0.078$), in line with the extended click effects in ID. Taken together, the click and intent-to-apply results trace a consistent within-job-seeker pattern that confirms and extends the main finding: in BM, and possibly also in ID, gender-fair titles increase both the probability of clicking and the intent to apply to a job ad.

For male job seekers, click and intent-to-apply results are reported in Tables A.41a through A.43b in the Online Appendix. Male click results at the dyadic level reveal a significantly negative effect in *IT & Development* ($p = 0.042$), indicating that gender-fair titles reduce male clicking in this category. However, this does not translate into any significant differences in ranking behavior (i.e., neither their propensity to rank a job ad, nor the assigned rank). In terms of consequential

outcomes, this is in line with finding no measurable male backlash in applications due to the usage of gender-fair titles.

Fixation results at the dyadic level are reported in Tables A.38a through A.39b in the Online Appendix; they corroborate the aggregate finding of a dissociation of shifts in visual attention and changes in application behavior, once again more consistent with a signaling explanation than one of novelty and visual salience.

4 Discussion

This paper studies whether a minimal intervention–switching job titles from the generic masculine to a gender-fair formulation–affects job-seeker behavior in naturally occurring labor market settings. In a large-scale randomized field experiment on a major German online job platform, we show that gender-fair titles substantially increase the number of female applications in BM, with no decline in male applications. This represents an increase of roughly 50% in female applications. The effect is mirrored by a significant increase in female clicks, suggesting that the treatment already operates at the stage of initial consideration: gender-fair titles shift which job ads enter female job seekers' consideration sets, and these early-stage effects translate into higher application numbers. In contrast, we find no comparable effects in ID and MS.

The pattern of heterogeneity is consistent with a signaling interpretation. In our framework, job titles serve as low-cost signals that update job seekers' beliefs about organizational culture. Signals matter most when prior beliefs are sufficiently elastic. In BM, where the gender composition is more balanced, a gender-fair job title plausibly shifts beliefs about workplace inclusiveness. In strongly male-dominated domains such as ID, prior beliefs may be too extreme for such a minimal intervention to meaningfully update expectations. In MS, the gender-fair title might not convey additional information to update.

Further evidence supports this mechanism. Within ID, we observe positive effects on female applications for non-IT companies posting ID jobs, but not for IT companies. This distinction is informative: when the company's broader identity is not tightly associated with a male-dominated environment, the gender-fair title becomes a more credible signal, aligning closely with a Bayesian updating interpretation.

We complement the field experiment with two mechanism studies. First, a study with hiring experts reveals that the decision to use gender-fair titles is driven more by genuine support for gender equality than by strategic beliefs about their effectiveness. Although hiring experts anticipate

both a female boost and a male backlash, the latter is not supported by our field data. Because gender-fair title usage stems from underlying attitudes rather than mere strategic concerns, job seekers can plausibly interpret it as an authentic signal of the employer's type.

Second, a laboratory experiment with job seekers replicates the category-specific field pattern. Gender-fair titles significantly increase female clicks in BM. This translates to the individual level, where positive effects on female clicking and intent to apply extend to ID as well, implying a broader reach of the signal when analyzed with greater statistical power. The decrease in male clicks in the lab data in ID does not translate to differences in the intent to apply, warranting some caution when using gender-fair titles in strongly male-dominated job categories. Yet the lab data mirrors the absence of a consequential male backlash in the field.

Eye-tracking evidence further refines the mechanism. The treatment differences in clicks and intent to apply in ID and BM do not coincide with systematic shifts in the relative share of fixations directed at gender-fair title regions, neither overall nor for female job seekers. This pattern is consistent with belief-based evaluation rather than attentional capture through novelty.

Taken together, the evidence reveals that gender-fair job titles can function as a low-cost signal that expands the applicant pool when prior beliefs have room for updating. The effectiveness of such a signal depends crucially on the context–specifically on baseline beliefs and the informational environment in which the signal is embedded.

Our signaling interpretation aligns with the concurrent findings of Del Carpio and Fujiwara (2026), who study gender-neutral language in job ads on a Spanish-language tech platform. Their key result–that treatment effects appear only when few neighboring ads are gender-neutral–demonstrates that the informational value of the signal diminishes as it becomes widespread. This spillover pattern is conceptually analogous to our cross-category heterogeneity: in both cases, the signal is effective precisely where it stands out against baseline expectations. Their complementary survey experiment confirms that female applicants update beliefs about company culture in response to the language, mirroring the belief-based mechanism supported by our eye-tracking evidence. Together, the two studies suggest that the signaling role of gender-fair language generalizes across languages (German, Spanish), institutional settings, and intervention intensities–from editing entire ads to changing just the title.

These findings carry practical implications. In industries facing acute labor shortages, gender-fair language can help employers tap into a broader talent pool at virtually no cost. Our results suggest that such language should be viewed not only as a matter of social inclusion but also as one of

economic efficiency: by removing subtle barriers that discourage otherwise qualified applicants, gender-fair titles can enhance labor-market matching. The sizable magnitude of our treatment effect–roughly 50% more female applications in BM–combined with the widespread but incorrect belief among hiring experts in a male backlash suggests that gender-fair job titles are currently underused.

Several limitations warrant acknowledgment. Our intervention pertains to a specific gender-fair form (the gender colon) in German, a strongly gendered language; effects may differ with other formulations or in other linguistic contexts. While we identify clear effects on application behavior, we cannot trace how additional applicants fare further along the hiring process. Finally, as societies become more accustomed to inclusive formulations, their signaling value may evolve. Despite these caveats, by combining large-scale field evidence with complementary mechanism studies, this paper provides unusually direct evidence that small, low-cost signals can meaningfully alter labor market behavior when beliefs are sufficiently responsive.

5 Conclusion

In an RCT conducted on a major German job platform, we tested the impact of using gender-fair job titles on application numbers by both men and women in general and for women specifically. Our findings indicate that gender-fair titles positively influenced applications in the *Business & Management* category but did not significantly affect *IT & Development* or *Marketing & Sales*. A possible driver for increased engagement appears to be the gender-fair titles attracting more attention. Within the *IT & Development* category, we find a positive effect for positions advertised by non-IT companies, suggesting that jobs in the male-dominated IT industry may require more substantial interventions to shift (particularly female) applicant behavior.

Complementing the field evidence, an online experiment with hiring experts shows that gender-fair titles are perceived not merely as an instrumental means to attract more applicants but actually reflect the employers' support of gender diversity. Both male and female hiring experts expect an increase in female applications, but also anticipate a possible "male backlash" in the number of applications–a perception that is inconsistent with our field data. A corresponding laboratory study with job seekers further reveals that gender-fair titles increase female clicks and intent to apply for *Business & Management* and *IT & Development* jobs but not for *Marketing & Sales*, mirroring the field evidence.

Taken together, our findings provide insights into how gender-fair job titles affect the volume

and gender breakdown of applications and highlight the need for context-specific evaluation of gender equality interventions. Despite our three-character intervention of adding ":in" to the job titles in the treatment being a minimal linguistic change, it demonstrates a substantive potential to attract more female applicants and therefore affect gender segregation at an early stage of the hiring process. Relative to the generic masculine, the use of gender-fair titles led to an increase of about 50% in female applications in *Business & Management* jobs.

The heterogeneity across job categories underscores the importance of aligning interventions with industry-specific norms and gender compositions. Gender-fair language should thus be viewed as one element in a broader strategy to reduce structural barriers to gender equality in labor markets. Understanding how such language-based interventions interact with employer perceptions, job seeker identities, and long-run hiring outcomes remains a promising direction for future research. More broadly, our study contributes to the growing literature in economics examining how language influences decision-making and economic behavior, offering evidence that linguistic framing can serve as a low-cost, effective tool to foster a more inclusive workforce within organizations and a more efficient labor market.

References

ABRAHAM, L., J. HALLERMEIER, AND A. STEIN (2024): "Words matter: Experimental evidence from job applications," *Journal of Economic Behavior & Organization*, 225, 348–391.

BANFI, S. AND B. VILLENA-ROLDAN (2019): "Do high-wage jobs attract more applicants? Directed search evidence from the online labor market," *Journal of Labor Economics*, 37, 715–746.

BARRON, J. M. AND J. BISHOP (1985): "Extensive search, intensive search, and hiring costs: New evidence on employer hiring activity," *Economic Inquiry*, 23, 363–382.

BLAU, F. D. AND L. M. KAHN (2017): "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, 55, 789–865.

BUNDESAGENTUR FÜR ARBEIT (2021): *Klassifikation der Berufe 2010 – überarbeitete Fassung 2020, Band 1: Systematischer und alphabetischer Teil mit Erläuterungen*, Nürnberg: Bundesagentur für Arbeit, November 2021.

CAMERON, A. C. AND P. K. TRIVEDI (1990): "Regression-based tests for overdispersion in the Poisson model," *Journal of econometrics*, 46, 347–364.

CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): "oTree–An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

COFFMAN, K. B., M. R. COLLIS, AND L. KULKARNI (2024): "Whether to apply," *Management Science*, 70, 4649–4669.

CORTÉS, P., J. PAN, L. PILOSSOPH, E. REUBEN, AND B. ZAFAR (2023): "Gender differences in job search and the earnings gap: Evidence from the field and lab," *The Quarterly Journal of Economics*, 138, 2069–2126.

DEL CARPIO, L. AND T. FUJIWARA (2026): "Spillovers and Scalability in Job Ad Experiments: Evidence from Gender-Neutral Language," Working Paper.

DELLAVIGNA, S. AND D. POPE (2018): "What motivates effort? Evidence and expert forecasts," *The Review of Economic Studies*, 85, 1029–1069.

DUTZ, R., S. HUBNER, AND C. PEUS (2022): "When agency "fits" regardless of gender: Perceptions of applicant fit when job and organization signal male stereotypes," *Personnel Psychology*, 75, 441–483.

DUTZ, R., S. REHBOCK, AND C. PEUS (2020): "FührMINT Gender Decoder: Subtile Geschlechtskodierung in Stellenanzeigen erkennen und auflösen," *Personal in Hochschule und Wissenschaft entwickeln*, 5, 87–95.

EICKER, F. (1967): "Limit theorems for regressions with unequal and dependent errors," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Berkeley, CA: University of California Press, 1, 59–82.

FLUCHTMANN, J., A. M. GLENNY, N. A. HARMON, AND J. MAIBOM (2024): "The Gender Application Gap: Do men and women apply for the same jobs?" *American Economic Journal: Economic Policy*, 16, 182–219.

GANESAN, M., S. P. ANTONY, AND E. P. GEORGE (2018): "Dimensions of job advertisement as signals for achieving job seeker's application intention," *Journal of Management Development*, 37, 425–438.

GAUCHER, D., J. FRIESEN, AND A. C. KAY (2011): "Evidence that gendered wording in job advertisements exists and sustains gender inequality." *Journal of Personality and Social Psychology*, 101, 109.

GOLDIN, C. (2014): "A grand gender convergence: Its last chapter," *American Economic Review*, 104, 1091–1119.

GORNY, P. M. AND P. NIEKEN (2023): "Gender-Fair Framing of Job Titles." AEA RCT Registry. February 22. DOI: https://doi.org/10.1257/rct.8060-4.0.

GREINER, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

HSIEH, C.-T., E. HURST, C. I. JONES, AND P. J. KLENOW (2019): "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 87, 1439–1474.

HUBER, P. J. ET AL. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, 1, 221–233.

INSTITUT FÜR ARBEITSMARKT- UND BERUFSFORSCHUNG (IAB) (2025): Last accessed: Nov 2025.

JAKIELA, P. AND O. OZIER (2020): "Gendered language," IZA Discussion Papers.

LOEWENSTEIN, G. AND Z. WOJTOWICZ (2025): "The economics of attention," *Journal of Economic Literature*, 63, 1038–1089.

MAS, A. AND A. PALLAIS (2017): "Valuing alternative work arrangements," *American Economic Review*, 107, 3722–3759.

MUMMOLO, J. AND E. PETERSON (2019): "Demand effects in survey experiments: An empirical assessment," *American Political Science Review*, 113, 517–529.

OPENAI ET AL. (2023): "GPT-4 Technical Report," ArXiv preprint 2303.08774.

PAN, J. (2015): "Gender segregation in occupations: The role of tipping and social interactions," *Journal of Labor Economics*, 33, 365–408.

RAFAELI, A. (2000): "Projecting an organizational identity: Lessons from employment advertisements," *Corporate Reputation Review*, 3, 218–239.

WHITE, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817–838.

ZIZZO, D. J. (2010): "Experimenter demand effects in economic experiments," *Experimental Economics*, 13, 75–98.

A A Model of Organizational Culture Signaling

We adapt the willingness-to-apply framework in Coffman et al. (2024) to a setting in which job-ad language signals organizational culture rather than clarifying job qualifications. Consider a job seeker who decides whether to *consider* a vacancy (see, click, read, apply), $a \in \{0, 1\}$; not considering yields outside option A. Considering entails cost $c \geq 0$ and yields a payoff that depends on the workplace's organizational culture.[26]

Each vacancy is posted by an employer with an underlying culture type $\theta \in \{I, T\}$, where I denotes an *inclusive* culture and T a *traditional* culture.[27] Job seekers have gender type $\tau \in \{F, M\}$ and differ in their valuation of culture. Let the incremental value of an inclusive (rather than traditional) culture be

$$g_\tau \equiv V_\tau(I) - V_\tau(T).$$

To interpret this in terms of typical empirical findings in the literature, one would associate a "male backlash" with $g_M < 0$, whereas $g_F > 0$ would indicate that women expect higher utility under an inclusive organizational culture.

Vacancies are grouped into categories indexed by C (e.g., industries or occupations) that differ in baseline gender composition and related expectations about workplace culture. Let

$$p_C \equiv \Pr(\theta = I \mid C)$$

denote the prior probability, common among job seekers, that a vacancy in category C comes from an inclusive culture. Intuitively, p_C is low in extremely male-dominated categories, intermediate in mixed categories, and high in categories that are perceived as already inclusive.

A vacancy title contains a language signal $s \in \{GF, B\}$, where GF stands for gender-fair language and B for baseline (non-gender-fair) language. Signal realizations are informative about θ because employer types differ in their intrinsic incentives to use gender-fair language. Inclusive employers weakly prefer using GF–reflecting internal norms or organizational practices–whereas traditional employers incur a disutility from doing so. Consequently, inclusive employers use GF with (weakly) higher probability, while traditional employers may nevertheless adopt GF strate-

[26] You can think of this as the sum of expected utility flow from the time span working in this position, given (and thus also abstracting from) a fixed probability of eventually being hired.

[27] This is a reduced representation of workplace culture intended to capture perceived inclusivity/belonging, not productivity or qualification.

gically when the perceived benefits in attracting applicants outweigh this disutility.[28]

We keep the employer side deliberately simple and represent equilibrium signal frequencies directly:

$$\Pr(s = GF \mid \theta = I) = \alpha,$$
$$\Pr(s = GF \mid \theta = T, C) = \beta_C,$$

with

$$0 \leq \beta_C < \alpha \leq 1.$$

Hence

$$\Pr(s = B \mid \theta = I) = 1 - \alpha,$$
$$\Pr(s = B \mid \theta = T, C) = 1 - \beta_C.$$

We treat α as category-invariant for simplicity. The idea is that the use of gender-fair language reflects an underlying organizational norm that is largely orthogonal to the occupational category itself. Allowing α to vary with C would mainly shift baseline signal frequencies without affecting the belief-updating logic, so fixing α keeps the notation parsimonious without loss of generality for our purposes. The category index on β_C captures variation in imitation incentives across job categories. In our setting, gender-fair language is the relatively new signal, whereas baseline wording represents the historical default. As a result, traditional employers may adopt gender-fair language strategically to appeal to applicants even when the underlying workplace culture is unchanged. The scope for such "pinkwashing" plausibly differs across categories–for instance, depending on applicant expectations or the reputational value of appearing inclusive–so we allow the imitation probability β_C to vary with C. By contrast, we abstract from inclusive employers mimicking traditional language, which would correspond to category-specific variation in α. This asymmetry reflects that baseline language is the default in the absence of signaling incentives, while the strategic margin arises primarily from adopting the newer signal.

[28] One can formalize this intuition in a standard signaling framework in which employer types choose the language signal s to maximize expected hiring payoffs net of type-specific signaling costs. In such a model, inclusive employers have weakly lower costs of using GF, whereas traditional employers face strictly higher costs but may mimic when applicant responses make doing so profitable. We abstract from modeling the employer's signaling decision explicitly because our empirical focus is on job seekers' belief updating and application decisions.

Upon observing (C, s), seekers form a posterior belief

$$q_C(s) \equiv \Pr(\theta = I \mid C, s).$$

By Bayes' rule, for any signal realization with positive probability,

$$\begin{aligned} q_C(GF) &= \frac{\alpha\, p_C}{\alpha\, p_C + \beta_C(1 - p_C)}, \\ q_C(B) &= \frac{(1-\alpha)\, p_C}{(1-\alpha)\, p_C + (1-\beta_C)(1-p_C)}. \end{aligned}$$

Thus, belief updating is governed by (i) priors p_C (baseline expectations in category C) and (ii) credibility (the gap between α and β_C).

Let U denote the culture-independent expected net utility from considering the vacancy, relative to not considering (outside option A), and net of costs:

$$U \equiv \underbrace{\mathbb{E}[\text{material payoff} \mid a = 1]}_{\text{expected material payoff from considering}} - A - c.$$

The total expected utility $\mathbb{U}$ from considering after observing (C, s) is

$$\begin{aligned} \mathbb{U} =& U + q_c(s) V_\tau(I) + (1 - q_c(s))\, V_\tau(T) \\ =& U + q_c(s) V(T) + q_c(s)\, [V_\tau(I) - V_\tau(T)] \\ =& U + V_\tau(T) + g_\tau q_C(s), \qquad\qquad g_\tau \equiv V_\tau(I) - V_\tau(T). \end{aligned}$$

The seeker considers the vacancy whenever this quantity is nonnegative.

Equivalently, the decision can be written as a posterior-threshold rule. Let

$$\bar{q}_\tau \equiv -\frac{U + V_\tau(T)}{g_\tau}.$$

Then, the decisionto consider a, follows the rule

$$a = \begin{cases} 1 & \text{if} \quad q_C(s) \geq \bar{q}_\tau \ \wedge . \ g_\tau > 0, \\ 1 & \text{if} \quad q_C(s) \leq \bar{q}_\tau \ \wedge . \ g_\tau < 0, \\ 1 & \text{if} \quad U + V_\tau(T) \geq 0 \ \wedge . \ g_\tau = 0, \\ 0 & \text{otherwise.} \end{cases}$$

A gender gap in consideration arises whenever the decision indices differ across τ. In particular, differences in g_τ and/or in $V_\tau(T)$ can generate different consideration behavior for women and men at the same (C, s).

The treatment effect of switching a title from B to GF for type τ is

$$TE = \begin{cases} > 0 & \text{if } \ q_C(B) < \bar{q}_\tau \leq q_C(GF) \ \text{ and } \ g_\tau > 0, \\ < 0 & \text{if } \ q_C(B) \leq \bar{q}_\tau < q_C(GF) \ \text{ and } \ g_\tau < 0, \\ = 0 & \text{otherwise.} \end{cases}$$

This implies that effects need not be monotone in the baseline category expectations p_C. For women, if $g_F > 0$, then when p_C is very low, even a gender-fair title may not raise $q_C(GF)$ enough to exceed $\bar{q}_F$; moreover, when credibility is low (i.e. β_C close to α), $q_C(GF) \approx q_C(B)$ and updating is small. When p_C is very high, $q_C(B)$ may already exceed $\bar{q}_F$, leaving little scope for additional changes in behavior. By contrast, for intermediate p_C–where inclusive culture is plausible but uncertain–a switch from B to GF can move $q_C(s)$ from below to above $\bar{q}_F$, generating the largest increase in female consideration.

Finally, finding "no male backlash" across job categories would imply $g_M = 0$ within our model, in which case changes in $q_C(s)$ do not affect men's decisions.

B Estimation Strategy

We conduct Tobit regressions for the number of applications (overall and by gender) to account for left-censored job ads that did not receive a single application.[29] We conducted our regressions in five steps.

[29] We started by fitting Poisson models and tested for overdispersion. Using a Pearson goodness-of-fit test (Cameron and Trivedi, 1990), we can reject this specification with $p < 0.001$. Negative binomial regression results are qualitatively identical and results can be found in Tables A.17a through A.18e in the Online Appendix.

First, we introduced only the treatment dummy *Gender-Fair* to report our raw treatment effect across the categories. Then, we added interaction terms for *Gender-Fair* and *ID*, as well as for *Gender-Fair* and *MS*, respectively, to disentangle the treatment effects in each job category. Note that this implies that *BM* serves as a statistical baseline. In the next step, we included information on the company advertising the job, such as the *Company industry* and *No. of employees*. Then, we included controls for the application procedure, such as *No. of required documents* and *No. of skills*. In the last step, we added various measures obtained from analyzing the job ad text, such as *No. of words* and *Informal you*. We report the coefficients for the main effect and interactions with the categories in the paper. Full tables, including the coefficients for all controls, are reported in Tables A.8a through A.9e in the Online Appendix.

Although our preregistered analysis primarily focuses on treatment differences within job categories, we conclude our analysis by reporting the pooled treatment effects for completeness.

Throughout our regression analysis, we use heteroskedasticity-robust sandwich estimators (Eicker, 1967; Huber et al., 1967; White, 1980) to account for potential problems of inflated variance and heteroskedasticity more generally.

C Description of Further Controls in the RCT Data

C.1 Company and Application-related Variables

Many job categories are different from the advertising company's industry. For example, if an IT company advertised a job as an HR officer, *BM* would be equal to one, as the job is in the category *Business & Management*. Therefore, the variable *Company industry* indicates the advertising company's industry.[30] Each company's size is categorized into six employee-number ranges and stored in *No. of employees*. The categories are "1-24," "25-49," "50-99," "100-249," "250-499," and "500+." We have the five-digit ZIP codes for companies. Because, in some instances, this makes companies individually identifiable, and because groups are too small for a reasonable analysis, we coded *First zip-code digit* to group all job ads with the same first digit of the zip-code to cover larger regions.

There are two key metrics for how demanding the application process is for each applicant. The variable *No. of required documents* represents the number of documents required for an applicant's application to the advertised job. Furthermore, job ads can include information on the required skills for a job. For each skill, the level of that skill is stated on a scale from one to four. We have

[30] A complete list of these values can be found in Table A.1 in the Online Appendix.

each skill as a dummy for whether it is requested in a job ad or not, as well as the corresponding skill levels. Again, since some skills are only requested in very few or even only single job ads, we grouped these skills into 18 broader skill categories.[31] For each of these *Skill categories*, we have a dummy that is one if the skill is requested and zero otherwise. Beyond these dummies, the number of skills required for an ad is given by *No. of skills*, and the average level of the skills required in a job ad is given by *Avg. level of skills*.

C.2 Text-Mining and LLM Classification

In addition to the quantitative data, we also have the text of each job ad. The information contained in these texts is crucial to understanding the potential mechanism behind our treatment effects. We thus incorporate it into our data analysis by using NLP techniques.

In the most conservative approach, we use simple counts, like the job ad's *No. of words*. Similarly, we account for forms of gender-neutral markers in the job ad.[32] We account for the sum of the colon (":"), the asterisk ("*"), the underscore ("_"), and "Binnen-I" ("I") in *No. of gender-fair words*.

For some more contextual categorizations of the job ads, as opposed to finding and counting certain substrings of the job ad, we used the GPT4 API (OpenAI et al., 2023). We created a dummy equal to one if the job ad explicitly mentioned *Hybrid/remote* work or home office opportunities and a zero if it did not. Similarly, we created the variable *Informal you*, which is one, if a job ad is formulated using the informal you ("Du") and zero otherwise ("Sie").[33]

We use the word lists underlying the FührMINT Gender Decoder (Dutz et al., 2020) to count the *No. of agentic words* and *No. of communal words* in each job ad. On the one hand, agentic traits such as ambition, assertiveness, and competitiveness, often perceived as masculine, are generally associated more with men. This extends to words and phrases related to these characteristics. On the other hand, communal characteristics such as supportiveness, caregiving, and responsibility, which are often considered feminine, are typically attributed more to women. Again, this also applies to verbal expressions reflecting these traits. They have previously been shown to affect perceptions of job ads in other studies (see, e.g., Gaucher et al., 2011; Dutz et al., 2022).

[31] The list of these can be found in Table A.2 in the Online Appendix.

[32] Note that these are not due to our treatment as these are outside of the title. Whereas we could control the gender-neutrality of the title, we left the texts (job description, company information, etc.) unchanged.

[33] In German, there are two forms of addressing people. The informal ("Du") is used among friends, family, and younger people, denoting closeness. The formal ("Sie") is used for formal situations, strangers, and people in higher professional positions. This distinction resembles "thou" (informal) and "you" (formal) in historical English.

D Tables and Graphs with Further Summary Statistics and Results

D.1 Field Experimental Data

	ID	BM	MS	Total	
No. of Applications	14.87	11.72	14.95	13.73	$p < 0.001$
	(24.73)	(22.29)	(23.56)	(23.55)	
Company size category (employees)	3.605	3.471	3.259	3.455	$p < 0.001$
	(1.773)	(1.696)	(1.671)	(1.720)	
No. of required documents	1.705	1.621	1.710	1.675	$p = 0.109$
	(1.379)	(1.275)	(1.268)	(1.309)	
No. of required skills	6.369	5.291	5.379	5.684	$p < 0.001$
	(3.291)	(1.521)	(1.602)	(2.352)	
No. of words in job ad	225.5	238.2	242.6	235.1	$p < 0.001$
	(70.45)	(79.28)	(82.45)	(77.63)	
Hybrid/remote	0.671	0.578	0.634	0.626	$p < 0.001$
	(0.470)	(0.494)	(0.482)	(0.484)	

Note: The figures show means with standard deviations in parentheses. The p-values in the last column stem from a Kruskal-Wallis test of equality of population rank sums. Tables A.3 through A.5 in the Online Appendix contain the same summary statistics split by treatment.

Table 16: Summary statistics across job categories

The baseline use of gender-fair words also differs significantly across job categories (Kruskal-Wallis test, $p < 0.001$). On average, job ads contain 1.074 gender-fair words. This number is highest in category MS (1.379) and lower in BM (0.932) and ID (0.969).

Application rates likewise vary substantially across the job categories. Job ads in the BM, on average, received significantly fewer applications than ID and MS.

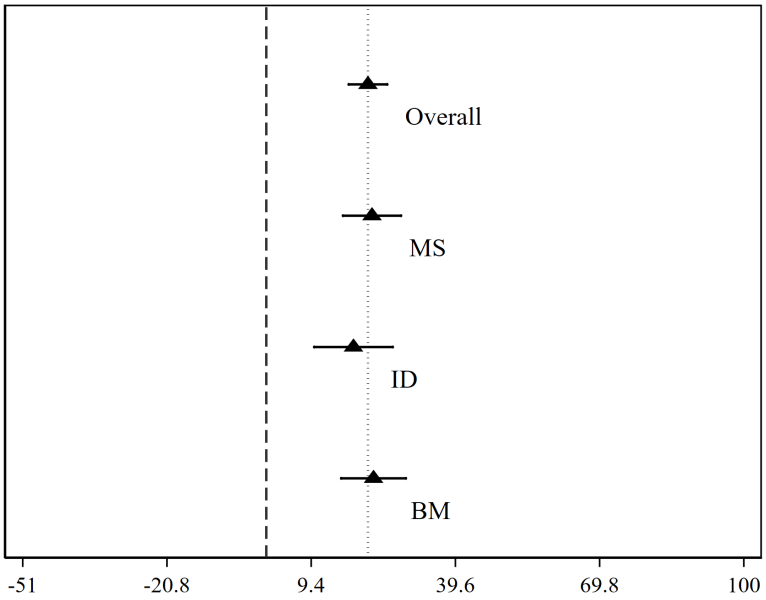
D.2 Hiring Expert Data

As shown in Table 17, hiring experts in Marketing & Sales (MS) more frequently selected gender-fair titles and text as well as the informal address, and selected salient team orientation, and flexible work arrangements similar to the pooled average, whereas hiring experts in Business & Management (BM) tended toward the opposite ends of these features. Hiring experts in IT & Development (ID) generally fell between these two groups, except for salient team orientation and flexible work, which they selected most frequently.

	ID	BM	MS	Total	
Informal	0.717	0.688	0.800	0.730	$p = 0.346$
	(0.455)	(0.466)	(0.404)	(0.445)	
GF Title	0.674	0.623	0.727	0.669	$p = 0.471$
	(0.474)	(0.488)	(0.449)	(0.472)	
GF Text	0.543	0.532	0.618	0.562	$p = 0.618$
	(0.504)	(0.502)	(0.490)	(0.498)	
Team	0.717	0.688	0.709	0.702	$p = 0.261$
	(0.455)	(0.466)	(0.458)	(0.459)	
Flex	0.826	0.714	0.818	0.775	$p = 0.934$
	(0.383)	(0.455)	(0.389)	(0.419)	

Note: The figures show means with standard deviations in parentheses. The p-values in the last column represent the results from Fisher exact tests.

Table 17: Job ad design configurations across job categories



Note: Each triangle shows the mean anticipated gender gap in the effect of gender-fair titles on applications, defined as the expected effect on female applications minus the expected effect on male applications (in percentage points on the belief elicitation scale). Whiskers denote 95% confidence intervals. Positive values indicate that hiring experts expect gender-fair titles to attract relatively more female than male applicants. The dashed and dotted lines mark zero and the overall mean, respectively.

Figure 2: Similar beliefs about the relative effect of gender-fair titles on male and female job seekers

D.3 Job Seeker Lab Data

D.3.1 Summary statistics

Table 18 reports summary statistics by job category as well as the pooled averages. Participants are, on average, 24 years old, with roughly half being female, and about one-third having a migration background. Kruskal-Wallis and Fisher exact tests reported in the final column indicate no statistically significant differences across groups, since participants were invited from the same university subject pool, and we explicitly balanced invitations by gender.

	ID	BM	MS	Total	
Age	23.72 (3.214)	23.77 (2.381)	26.18 (13.22)	24.35 (7.051)	$p_{KW} = 0.851$
Female	0.480 (0.505)	0.519 (0.505)	0.471 (0.507)	0.493 (0.502)	$p_F = 0.928$
Migration background	0.260 (0.443)	0.365 (0.486)	0.441 (0.504)	0.346 (0.477)	$p_F = 0.217$

Note: The figures show means with standard deviations in parentheses. The p-values in the last column represent the results from Kruskal-Wallis (KW) and Fisher exact tests (F).

Table 18: Summary of job seeker demographics

D.3.2 Preparation of Eye-Tracking Data

The raw eye-tracking data were preprocessed in several steps to obtain fixation counts per participant and area of interest (AOI). Each gaze sample contained left- and right-eye coordinates and a validity flag. Whenever both eyes were valid, gaze positions were averaged to obtain a single horizontal and vertical coordinate (g_x, g_y); if only one eye was valid, its position was used, and missing values were retained otherwise. Device timestamps, originally recorded in microseconds, were converted to milliseconds to allow for meaningful duration measures. Each sample was then assigned to a specific AOI, corresponding to a distinct element of the job description–for instance, the title, introductory text, task description, or compensation offer. On the *SearchPage*, participants could only see job titles, whereas on the *RankPage* all elements were potentially visible. Within each participant, page, and AOI, consecutive gaze samples recorded less than 12 ms apart were grouped into continuous "runs." For each run, start and end times were used to compute its duration. Runs exceeding at least 80 ms, 100 ms, or 150 ms were classified as fixations, providing alternative thresholds for robustness checks in line with standard practice. Finally, the number of fixations per participant and AOI was counted, yielding fixation-frequency tables that

summarize how intensively each section of the job advertisement attracted visual attention.