

Leader Identity, Evaluations, and Performance: Evidence from a Laboratory Experiment

Eric Wilken*

June 21, 2026

Abstract

Observers often use identity as a proxy for the characteristics they associate with effective leadership. It remains unclear whether these expectations of effectiveness influence how individuals behave when working under leaders of differing identities. This paper examines the gap between evaluations and behavior using a laboratory experiment¹ in a nonprofit, volunteer setting where motivation is non-pecuniary. Participants are randomly assigned to leader vignettes whose gender and sexual orientation vary. They are then asked to complete an effort-based data entry task and survey which aims to evaluate the effectiveness of their assigned leader. I find that women are rated as more effective and more representative of a typical leader, in the context of nonprofit volunteer coordination, than men. This advantage does not extend to gay women. Despite these differences in evaluations, there are no detectable differences in effort or accuracy across leader types. These findings provide suggestive evidence that stated evaluations of leadership may be a poor signal of actual productivity gains.

JEL Codes: D91, C91, M54, L30

Keywords: Leadership, Nonprofit Management, Identity Economics

*Department of Economics, University of Oregon. Email: ewilken@uoregon.edu Website: <https://ericwilken.com>

¹Prior to conducting the main analyses, the research design, hypotheses, outcome measures, and analytic plan were preregistered on the AsPredicted platform (AsPredicted #172,448). This study was approved under the exempt category by the University of Oregon Institutional Review Board Protocol #00001141.

1 Introduction

The stereotypical leader is often described as "ambitious, self-sufficient, and dominant" (Powell et al., 2002; Schein, 2001), and as possessing masculine traits or physical features (Giacomin et al., 2022; Koenig et al., 2011). These descriptions reflect widely held beliefs about what makes an effective leader and reflect the image of a leader as consistently portrayed in media from Machiavelli's *The Prince* to *The West Wing*. However, it remains unclear whether these stated and portrayed preferences correspond to how individuals truly respond to leader identity.

Most people do not fit this prototypical image, whether due to personality, appearance, or social identity. Prior research shows that women and members of sexual minority groups are often perceived as less prototypical and less desired as leaders (Eagly & Chin, 2010; Liberman & Golom, 2015; Pellegrini et al., 2020). These perceptions are typically measured through stated evaluations, such as survey responses or hypothetical judgments. Yet, such measures may not reflect how followers behave when asked to exert effort on behalf of their leader.

This distinction between stated and revealed preferences is central to understanding leadership effectiveness. If individuals report preferences for certain leader characteristics but do not act on those preferences in real tasks, then commonly used measures of leadership perception may provide a misleading guide to actual performance. This raises the possibility that leadership positions are allocated based on beliefs that do not translate into differences in productivity.

I propose a laboratory experiment to examine whether stated preferences over leader characteristics, specifically social identity, align with revealed measures of follower effort. The study focuses on leadership in the context of nonprofit organizations and volunteerism, where financial incentives are limited and motivation must come from non-pecuniary sources. In this setting, leadership plays a particularly important role in motivating participation.

To test the relationship between stated and revealed preferences, participants are ran-

domly assigned to a fictitious leader whose identity varies along observable dimensions, including gender and sexual orientation. These identity characteristics are used as a tractable way to introduce variation in leader perceptions, as prior research suggests they are associated with systematic differences in stated evaluations. Participants then complete an effort-based task and report their attitudes toward the assigned leader.² This design allows for a direct comparison between how individuals say they evaluate leaders and how they behave when working under them.

I find that leader identity shapes how followers evaluate their leaders but has no detectable effect on how hard they work. In the context of nonprofit volunteer coordination, women leaders are rated as more effective and more representative of a typical leader than men. This advantage does not extend to gay women, who are viewed as less prototypical despite the generally favorable ratings received by women leaders overall. When it comes to actual task performance, there are no meaningful differences across any of the four leader types. Followers worked just as hard and just as carefully regardless of whether their assigned leader was a straight man, a straight woman, a gay man, or a gay woman. This disconnect between stated evaluations and observed behavior raises important questions about whether the stated assessments organizations use to select and promote leaders are optimal.

2 Background

Leadership is not an innate trait but rather a process in which leaders and followers work together toward a shared goal within a particular social and cultural context (McManus & Perruci, 2020). Psychologists, sociologists, and leadership scholars have long studied leadership as a “relational process” (Bass, 1985; Fischer et al., 2017; Gardner et al., 2011; Hooper, 2016; Sheer, 2015), which means that leadership emerges through social interaction and contextual dynamics, even when the “leader” is a title granted by fiat. In contrast,

²Simple effort-based tasks, such as data entry or counting tasks, are commonly used in experimental economics to measure how incentives influence motivation (Charness et al., 2018).

economists and management scholars often focus on how the behavior and strategies of leaders influences task outcomes (Antonakis et al., 2022; Aunga & Masare, 2017; Grossman & Hart, 1992; Kim, 2002; Lazear, 2012; Puni et al., 2014). Within this framework, leadership is frequently conceptualized as a position that motivates subordinate effort through reward and punishment mechanisms.

Historically, these two approaches have remained largely separate. Garretsen, Stoker, and Weber (2020) argue that integrating these perspectives could benefit leadership research. They propose developing a unified theory of leadership that reduces conceptual ambiguity, more frequently identifies causal relationships, and investigates leadership beyond purely individual or group-level analysis. In this paper, I take this approach –combining elements of identity economics, task-motivation, and social perceptions of leadership– to consider the experience of gay and lesbian leaders.

In their foundational work on identity economics, Akerlof and Kranton, 2000 argue that individuals’ preferences are influenced by social, cultural, and personal norms. These norms shape individual identity and affect how individuals interact with economic institutions. Akerlof and Kranton later apply this framework to firms, suggesting that organizations operate more efficiently when employees feel that they share a common identity within the organization (Akerlof & Kranton, 2005). This implies that firms may benefit from investing resources that foster a sense of belonging among employees.

How do leaders do this? There is extensive work that considers the taxonomy and effectiveness of leadership styles. Numerous leadership styles have been theorized, including charismatic leadership and servant leadership (Antonakis et al., 2022; Greenleaf & Spears, 1998). One concept particularly relevant to this study is *G Quotient Leadership*, a theory developed through research on successful gay executives (Snyder, 2006). This theory suggests that there may be systematic differences in leadership styles between straight and gay leaders. While leadership styles are not inherently restricted to particular groups, some scholars argue that certain groups may be more likely to adopt particular styles. Researchers have examined

how different leadership styles affect performance both within and outside of firms (Aunga & Masare, 2017; Meslec et al., 2020; Puni et al., 2014). In contrast, the present study holds leadership style constant while varying leader identity. This design allows for a clearer examination of the relationship between leader identity and follower behavior in a controlled setting. To the degree successful leaders create a sense of common identity which affects subordinate performance and leadership perceptions, I examine the role of leader identity, independent of style, directly linking the two threads of Akerlof and Kranton work.

2.1 Leader identity

Research on leadership perception has increasingly focused on how identity-based biases shape who is recognized as a legitimate leader and who ultimately gains access to leadership roles. Much of this work has centered on gender, offering substantial insight into how men and women are evaluated differently. In contrast, the role of sexual orientation in shaping leadership perceptions remains comparatively underexplored.

Two of the most common terms one reads about women and leadership are the “glass ceiling” and the “glass cliff”. The glass ceiling is the term used to describe the persistent difficulty women face in reaching top leadership roles despite comparable qualifications. Similarly, the glass cliff highlights the common occurrence of when women are promoted, they are disproportionately placed into leadership positions during periods of crisis or heightened risk (Cotter et al., 2001; Ryan & Haslam, 2005; Ryan et al., 2016). Evidence suggests that women are less likely to be selected for stable, high-reward leadership roles but more likely to be appointed when the probability of failure is high, exposing them to greater scrutiny and a higher likelihood of negative evaluations (Ryan & Haslam, 2005; Ryan et al., 2016). These dynamics complicate standard interpretations of leadership outcomes, as observed differences in performance may reflect differences in the environments in which leaders are placed rather than differences in underlying ability.

Work on the perceptions of women leaders shows that they are shaped by a mismatch

between gender norms and leadership expectations. Role congruity theory in psychology predicts that women are evaluated less favorably in leadership roles because these roles emphasize traits such as dominance and assertiveness that conflict with traditional expectations of femininity (Eagly & Karau, 2002; Heilman, 2001; Ridgeway, 2001). As a result, women may be perceived as less effective leaders even when their behavior is identical to that of men (Eagly & Chin, 2010). These perceptions are reinforced by widely held beliefs about what leaders “look like.” Across multiple research paradigms, leadership traits are consistently associated with masculinity, suggesting that the prototypical leader is implicitly male (Koenig et al., 2011; Powell et al., 2002; Schein, 2001). This pattern persists across contexts and contributes to systematic differences in how men and women are evaluated as potential leaders (Giacomin et al., 2022).

However, when evaluating the performance of leaders rather than leadership perceptions, the evidence is more mixed. A large body of work finds that women are rated as equally or slightly more effective leaders than men across a wide range of settings (Eagly et al., 2003; Paustian-Underdahl et al., 2014). This divergence between perceived leadership and leader effectiveness highlights a gap between beliefs about leadership and beliefs about measurable outcomes. Gender may operate as a diffuse “status characteristic”, shaping both selection into leadership and perceptions of those in those roles (Correll, 2004; Ridgeway, 2001) without needing reinforcement of differential effectiveness. These dynamics can lead to inefficient allocation of leadership positions if evaluators rely on biased priors rather than observed performance.

In the case of gay and lesbian leaders, discrimination based on sexual orientation provides one possible explanation for why such biased perceptions could exist. Research suggests that some of the perceived ineffectiveness of queer leaders is explained by evaluators’ levels of prejudice, yet research suggests that something more than discrimination—either implicit or explicit—is at play (Morton, 2017). These perceptions also appear to manifest in the real world in the hiring of leadership positions: queer leaders receive fewer formal leadership roles

than their straight counterparts, even when they are perceived as equally competent (Aksoy et al., 2019; Fasoli et al., 2017; Frank, 2006; Niedlich & Steffens, 2015). Often referred to as the “lavender ceiling,” the scale of this disparity in leadership may be caused by a combination of in-group effects, stereotypes, and discrimination. According to the Williams Institute, in 2021 approximately 50.4% of LGBT employees reported not being out to their current supervisor and 25.8% were not out to any coworkers (Mallory et al., 2024). This likely reflects a strategic response to the lavender ceiling and the continuing presence of discrimination and anti-gay bias in the workplace.

Beyond the motivation of equitable access to leadership roles, discrimination is costly. When societies become more welcoming of queer individuals, higher GDP growth tends to follow, likely due to reduced transaction costs and reduced information loss associated with the avoidance of queer workers at the firm level (Badgett et al., 2019). There may therefore be additional gains from increased inclusivity in leadership if discrimination has led to an inefficient allocation of leadership positions.

3 Methods

The experiment addresses three main questions. First, I investigate whether leader identity, specifically gender and sexual orientation, influences the effort their followers exert when completing tasks. Second, I examine whether leader identity influences follower attitudes toward leaders in terms of leader typicality, and perceived success in other roles. Finally, I explore the link between followers’ beliefs about gender and sexuality and their stated preferences toward leaders.

3.1 Procedure

Participants completed the survey through the online platform Qualtrics. They began by reading and completing an informed consent form.

After consenting, participants were given instructions describing the study procedure and were randomly assigned to one of four fictitious leaders: (1) Gay Woman, (2) Gay Man, (3) Straight Woman, or (4) Straight Man. In all conditions, participants learned that the leader worked for a fictitious nonprofit organization whose mission was to improve adult literacy within the United States. The organization was identical across conditions.

Emily Reynolds is the Director of Outreach at the National Center for Literacy and Reading (NCLR). NCLR is a non-profit dedicated to eradicating poverty through expanding access to education in both rural and urban communities. Ms. Reynolds began working for NCLR in 2018 and currently leads a team of community organizers across the US to expand the offered literacy resources, such as free books and reading/writing classes. In addition to working at NCLR, Reynolds also sits on the board of directors of both the Kansas City Independence Day Parade, and Food for All, a non-profit aimed at reducing food insecurity and waste.

Reynolds was born in the bustling city of Kansas City, MO, nestled in the heart of the American Midwest. Growing up, Reynolds was deeply influenced by a strong sense of community and the Midwestern value of hard work. She has worked in the non-profit sector for more than a decade and has dedicated her career to fostering the expansion of opportunities in her community. Reynolds firmly believes that forging a more equitable world is best done at the local level and will require us to stand hand-in-hand with our neighbors.

Outside of work, Reynolds is an avid hiker and nature enthusiast. She often escapes with her husband to hike trails in the Ozarks or go sailing on Smithville Lake. Reynolds is also a voracious reader and has an extensive collection of books that span various genres, from classic literature to contemporary non-fiction. Regardless of the activity, Reynolds finds joy and balance in engaging in activities outside of her professional life with the people she cares about.

Once you have finished reading the passage, you may click the "Next Page" button to advance to the next screen and see a message from your leader.

Figure 1: Examples of the Leader Vignette

Participants then read a short vignette (see Figure 1) describing the assigned leader. The vignette resembled a short biography containing both professional and personal details about the leader. Participants then received a message from their leader describing the upcoming task. This message was designed to encourage effort based on previous literature on motivating workers (Antonakis et al., 2011, 2022; Fest et al., 2021; Meslec et al., 2020).

The task was introduced as a volunteer opportunity in which participants would enter publication information from books into a database to help the nonprofit catalog books. The task consists of two stages. The first stage involves training. Participants receive instructions describing how to identify information from a book's publication page (see Figure 2). Participants may spend as much or as little time on the training materials as they wish. The

amount of time spent on this training stage is recorded.

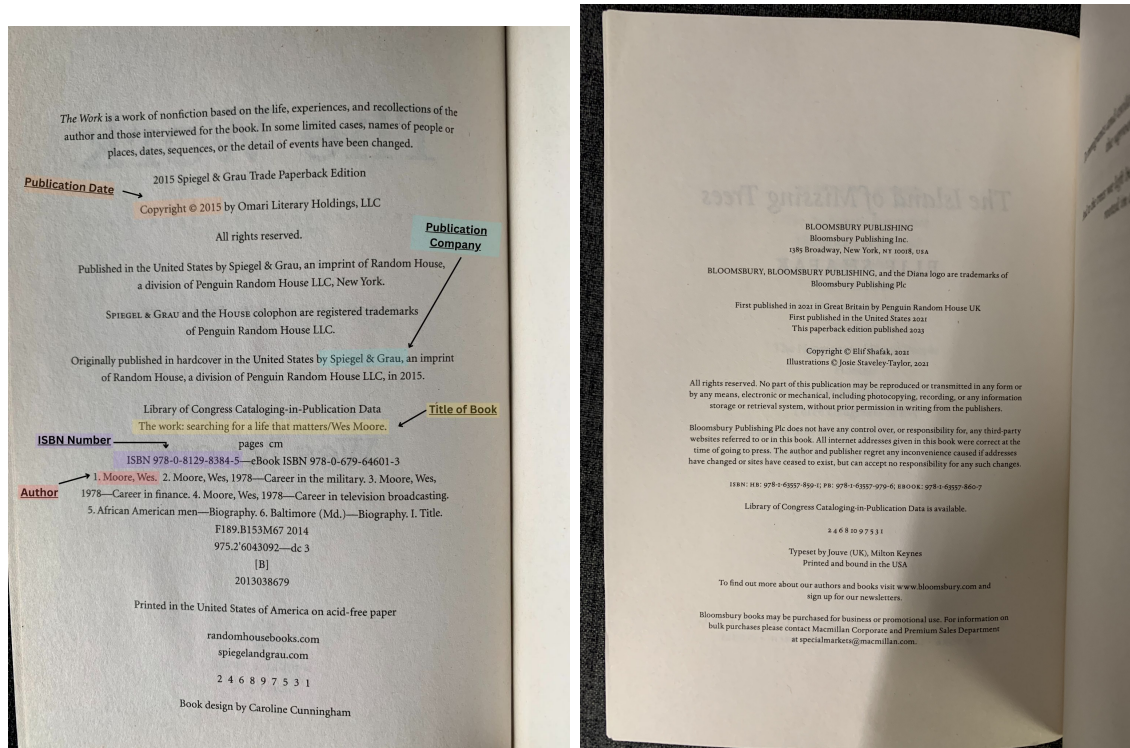


Figure 2: Examples of the Training and Data Entry Task

The second stage consists of a timed effort task. Participants are asked to enter information from as many book publication pages as possible within a 15-minute period. Each page must be displayed for a minimum of 17 seconds before participants can proceed to the next page, and there are enough pages available to ensure that participants remain engaged for the full duration of the task.

From this task, I collect several performance measures, including the time required to complete each entry, the accuracy of each entry, the total number of entries completed, the total number of entries completed correctly, and the percentage of entries completed correctly.

Participants concluded the study by completing a questionnaire containing the measures described below and demographic questions (age, gender, race, religion, and sexual orientation).

3.2 Measures

To test the effect leaders have on followers, I explore followers’ self-reported opinions of their assigned leader and their performance on the data-entry task. I highlight below each of the measures used in this experiment to capture underlying participant cultural beliefs, their beliefs about their assigned leader, and their revealed performance.

Data-Entry Task. In addition to the stated measure of leader effectiveness, I create several outcomes based on the data-entry task described above. These outcomes cover performance measures such as the number of entries attempted (simple effort-based outcome), the accuracy of entries (attention-based outcome), and an outcome that attempts to quantify the number of outcomes completed correctly.

One of the simplest forms of performance that we can measure is the total number of tasks completed by a volunteer. However, as a measure of effort, this requires that all tasks are completed at the same average level of quality. This is not the case. To capture the quality of each response, I must identify how similar a participant’s response is to the correct answer, where the correct answer is the information that is available to each participant on the publication pages.

Determining how similar a participant’s response is to the correct answer—or more generally how similar two text strings are to each other—is a common question in computational linguistics and computer science (Navarro, 2001). Edit distance is a way of quantifying how similar two strings are to each other by counting the minimum number of operations required to transform one string into the other (Navarro, 2001). I use the Jaro-Winkler similarity measure. The Jaro algorithm measures the number of characters in common between two strings while accounting for transpositions within the strings and is normalized between 0 and 1 (Winkler, 1990). A score of 0 means that the strings are identical and a score of 1 indicates that the strings are completely different. Winkler modified Jaro’s algorithm by making differences at the beginning of strings more costly than differences near the end of the

strings (Winkler, 1990). The Jaro-Winkler algorithm is typically used in the comparison of short strings such as names or, in the case of this paper, publishers, as it is computationally efficient compared to similar measures of edit distance.

Using this algorithm, I calculate the “correctness” of each participant response by transforming the output from the Jaro-Winkler algorithm to instead reflect a measure of similarity rather than distance. When determining what the correct answer should be, I construct and report three different levels of critique for correctness. I report these levels as “generous”, “normal”, and “strict.” Normal correctness is based off of a set of correct answers which require the participant to accurately input the correct title, author, and publisher (spelling matters but capitalization does not) shown on the publication page. However, participants are given flexibility on selecting the correct ISBN number. Strict correctness only permits the exact text string shown on the publication page, including capitalization and punctuation, as completely correct. Only the correct level of publisher and exact ISBN is accepted. Generous correctness allows for the participant to report either the parent publishing company or the specific imprint and be marked as correct.

Stated Leader Effectiveness and Typicality. Questions measuring leader effectiveness and typicality were adopted from Rosette and Livingston (2012). This measure has high internal consistency ($\alpha = 0.91$) and contains items that can be easily adapted to evaluate leadership effectiveness in this context. It includes four items measuring leader effectiveness and one item measuring leader typicality, which directly asks: “To what extent is NAME typical of a leader?”

Heteronormative Beliefs and Values. To measure participant heteronormativity, I adopt items from the heteronormative attitudes and beliefs scale developed by Habarth (2015). The original measure consists of 16 items divided into two subscales: the Essential Sex and Gender subscale and the Normative Behavior subscale. Because of concerns regarding participant attention when completing a 16-item instrument, I developed a short-form

adaptation of the original scale. The shortened measure contains five items—two from the Essential Sex and Gender subscale and three from the Normative Behavior subscale—and demonstrates acceptable internal consistency (Cronbach’s alpha $\alpha = 0.77$).

Other Measures of Beliefs To measure how prosocial a participant may be they are asked to report how many hours they have volunteered in the last 30 days. Additionally, participants are asked to rate their support of the nonprofit listed as the leaders place of work. These two variables are captured to verify that pro-sociality is balanced across conditions or control for it otherwise. These measures are used as controls.

4 Data

4.1 Participants

Participants were recruited from two university research pools: a marketing research pool and a psychology research pool.³ They received no monetary compensation for their participation in this research but did receive course credit. All participants had the option to opt out of the research at any time and could complete an alternative assignment for the required course credit.

Overall, 516 participants completed the survey. The final sample consisted of 62.1% women and 31.0% men. Participant age was relatively concentrated, with an average age of $M_{age} = 21.15$. Most participants identified as White (72.5%), and many identified as Christian (43.6%). In terms of sexual orientation, the sample includes participants who self-identified as straight (67.3%) and gay or lesbian (32.7%).

³Using a college sample for this study is primarily a matter of convenience. The ability to generalize findings from college subject pools to the broader population is limited (Hanel & Vione, 2016; Wild et al., 2022). This limitation is not a major concern for the present study. Understanding how a leader’s identity may influence the effectiveness of college volunteerism is itself an important question, as college students are frequently encouraged by their institutions to volunteer in their local communities. This volunteerism is meaningful for organizations working to accomplish their missions and visions. Additionally, identifying contexts in which leader preferences differ from general expectations allows us to comment more broadly on how strongly employers should rely on stated preferences when making hiring decisions.

The number of participants targeted for this study was informed by a statistical power analysis. Using a two-sided t-test with $\alpha = 0.05$ and statistical power of 0.8, I estimated that 504 participants would be required to detect a 0.25 standard deviation difference between in group means by leader gender or sexuality.

Table 1 presents participant demographic information separated by treatment condition. Several facets of participant identity varied across groups following random assignment. Any differences in underlying participant characteristics which appear across conditions are controlled for later in the analysis. Participants’ ability to correctly recall the leader’s identity ranged between a 53% and 82% success rate. The gay man leader condition was associated with the lowest rate of success in the manipulation check. Further examination indicates that this low success rate was primarily due to incorrect identification of the leader’s sexuality rather than the leader’s gender. In other words, participants often assumed that the gay man leader was straight.

Table 1: Participant Characteristics and Beliefs by Assigned Leader Condition

	Gay Man	Gay Woman	Straight Man	Straight Woman
Number of Participants	125	120	134	137
% Passed Attention Check	0.560	0.475	0.373	0.263
% Passed Manipulation Check	0.536	0.550	0.828	0.810
Mean Preference for Nonprofit	4.010	4.140	4.070	3.950
Mean Heteronormative Beliefs	2.240	2.300	2.340	2.280
Participant Volunteerism (Mean Monthly Hours)	12.100	16.600	12.200	11.600
Mean Participant Age	21.100	20.800	21.700	21.000
% Women	0.595	0.655	0.623	0.612
% White	0.700	0.716	0.729	0.752
% Straight	0.633	0.655	0.750	0.651
% Christian	0.450	0.447	0.458	0.391

4.2 Descriptive Statistics Across Groups

I calculate at the individual level the percentage of responses completed perfectly and the number of items completed correctly. The average of these values by experimental condition is shown in Table 2.

Table 2 provides a preliminary glance at differences in follower performance and stated opinions of leaders across leader identity conditions. Using the restrictive definition of cor-

rectness, differences across groups are modest. Followers assigned to straight woman leaders completed the most tasks completely correctly on average (21.5), while those assigned to gay man leaders completed slightly fewer (20.9). Accuracy under this definition is also similar across conditions, ranging from approximately 53.4% to 54.7%.

With a more relaxed definition of correctness, a similar pattern emerges. Followers assigned to straight woman leaders completed the most items on average (23.8), followed closely by those assigned to straight men (23.6) and gay women (23.5), while those assigned to gay men completed slightly fewer items (22.8). Overall accuracy is again quite similar across conditions, ranging from about 59.3% to 60.0%. Looking at maximum performance, followers assigned to straight woman leaders also achieved the highest number of items completed (32.3), with other groups performing similarly (31.4–32.2). Maximum accuracy is nearly identical across conditions, ranging from 79.6% to 80.6%. Taken together, these performance measures suggest only minor differences in productivity across leader identity conditions.

Stated opinions of leaders are also summarized in Table 2. The effectiveness and leader typicality measures are broadly similar across groups, though some modest differences emerge. Gay women leaders were rated most highly in terms of effectiveness, with an average score of 3.83 out of 5, while straight men received the lowest average rating (3.49). Leader typicality follows a similar pattern, with ratings ranging from 3.36 to 3.60 across conditions. These group averages provide limited evidence of systematic differences in how leaders are perceived across identity conditions.

Turning to the time data at the bottom of Table 2, there are differences in how participants interacted with the materials. The average training time is highest in the straight man leader condition (74.5 seconds), which appears to be driven by an outlier, likely representing a participant who left the experiment running before returning. Participants in the straight woman leader condition spent more time reading the vignette (75.1 seconds) relative to other groups. Excluding extreme observations reduces the average training time in the straight

Table 2: Performance, Leader Evaluations, and Time Use by Leader Identity Condition

	Gay Man	Gay Woman	Straight Man	Straight Woman
Number of Participants	125	120	134	137
<i>Performance (Strict Definition)</i>				
Items Completed Correctly	20.9 (8.20)	21.1 (8.26)	21.4 (7.99)	21.5 (7.88)
Percent Correct (%)	54.6 (14.2)	53.4 (15.7)	53.9 (14.0)	54.7 (14.9)
<i>Performance (Relaxed Definition)</i>				
Items Completed	22.8 (8.15)	23.5 (8.14)	23.6 (7.77)	23.8 (8.05)
Percent Correct (%)	59.3 (11.2)	59.3 (12.9)	59.6 (11.3)	60.0 (12.4)
<i>Maximum Performance</i>				
Max Items Completed	31.4 (11.4)	31.9 (11.2)	32.2 (10.8)	32.3 (11.6)
Max Percent Correct (%)	80.6 (11.9)	79.7 (14.1)	80.4 (12.8)	80.2 (14.1)
<i>Leader Evaluations</i>				
Leader Effectiveness (1–5)	3.61 (0.83)	3.82 (0.83)	3.49 (0.89)	3.61 (0.86)
Leader Typicality (1–5)	3.51 (0.76)	3.53 (0.88)	3.36 (0.79)	3.60 (0.81)
<i>Time Measures (seconds)</i>				
Training Time	47.6 (75.1)	39.2 (28.3)	74.5 (303.0)	36.9 (33.0)
Vignette Reading Time	85.7 (161.0)	62.7 (68.5)	59.1 (60.3)	75.1 (192.0)
Message Reading Time	33.9 (82.7)	25.9 (23.1)	23.8 (25.1)	25.9 (48.4)

*Entries are group means with standard deviations in parentheses.

man condition to approximately 67 seconds, suggesting that these differences are driven in part by a small number of unusually long durations.

5 Results

5.1 Leader Identity on Stated Beliefs and Revealed Performance

To explore the interaction between leader identity and follower behavior, I estimate a series of linear models where the outcome variables described above are regressed on experimental condition (leader gender, leader sexuality, and their interaction) indicator variables. The model takes the following form:

$$Y_i = \beta_0 + \beta_1 SEX_i + \beta_2 GEN_i + \beta_3 (SEX_i \times GEN_i) + \beta_4 X_i + \epsilon_i \quad (1)$$

where Y is the outcome variable of interest, SEX_i is a dummy variable representing the sexual orientation of the assigned leader for subject i (1 = gay, 0 = straight), GEN_i is a dummy variable representing assigned leader gender for subject i (1 = woman, 0 = man), and X_i is a vector of participant control variables. The comparison group captured by β_0 is a straight man leader.

Across all specifications, there are no statistically significant relationships between leader gender or sexual orientation and follower performance outcomes. This holds for both the number of items completed (Table 3) and measures of accuracy (Table 4), and is consistent across models with and without controls. The interaction between leader gender and sexuality is also small and imprecisely estimated in all cases.

That said, the magnitude and direction of several point estimates are informative. In Table 3, leaders who are women are associated with slightly higher levels of task completion across most specifications, though these effects are small and not statistically distinguishable from zero.

Leaders who are gay are generally associated with fewer items completed, particularly under the stricter definition of correctness, but again these estimates are imprecise. Taken together, these results suggest a weak and inconsistent relationship between leader identity and productivity, with no clear directional pattern.

Table 3: Effects of Leader Gender and Sexuality on Item Completion

	Items		Generous Items		Strict Items	
Intercept	23.649***	18.985***	32.231***	28.862***	21.410***	14.394**
	(0.672)	(5.138)	(0.937)	(6.677)	(0.691)	(5.517)
Leader is a Woman	0.117	0.064	0.097	-0.079	0.057	0.066
	(0.962)	(0.938)	(1.361)	(1.286)	(0.965)	(0.976)
Leader is Gay	-0.817	-0.348	-0.839	-0.006	-0.490	0.074
	(0.991)	(1.041)	(1.383)	(1.439)	(1.008)	(1.066)
Woman Leader $\times$ Gay Leader	0.576	0.093	0.453	-0.282	0.115	-0.449
	(1.417)	(1.356)	(1.982)	(1.875)	(1.427)	(1.397)
Controls	No	Yes	No	Yes	No	Yes
N	516	458	516	458	516	458

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

A similar story emerges for accuracy. In Table 4, the estimated effects of leader gender and sexuality fluctuate in sign across specifications and definitions of correctness. For example, female leaders are associated with slightly higher strict accuracy but slightly lower generous accuracy in some models, while gay leaders are associated with modest decreases in overall accuracy in certain specifications. However, these differences are small relative to baseline accuracy levels and are not statistically significant. This suggests that any relationship between leader identity and accuracy is, at best, limited and highly sensitive to how performance is measured.

Table 4: Effects of Leader Gender and Sexuality on Accuracy

	Pct. Correct		Generous Pct.		Strict Pct.	
Intercept	59.586***	29.158**	80.390***	46.938***	53.897***	20.189.
	(0.972)	(9.404)	(1.102)	(12.838)	(1.211)	(10.554)
Leader is a Woman	0.407	0.087	-0.171	-0.719	0.809	0.643
	(1.436)	(1.499)	(1.634)	(1.644)	(1.760)	(1.902)
Leader is Gay	-0.261	-0.917	0.182	-0.288	0.670	0.153
	(1.399)	(1.466)	(1.534)	(1.675)	(1.758)	(1.869)
Woman Leader $\times$ Gay Leader	-0.418	0.475	-0.750	-0.170	-1.970	-1.317
	(2.114)	(2.084)	(2.336)	(2.279)	(2.603)	(2.611)
Controls	No	Yes	No	Yes	No	Yes
N	516	458	516	458	516	458

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

In contrast, leader identity appears to play a more meaningful role in shaping follower evaluations (Table 5). In the controlled specifications, leaders who are women are associated with higher perceived effectiveness and are more likely to be viewed as typical leaders. The magnitude of these effects is modest but consistently positive, with some reaching conventional levels of statistical significance. By comparison, leader sexuality is not strongly associated with evaluations, as the corresponding coefficients are small and imprecisely estimated.

Table 5: Effects of Leader Gender and Sexuality on Leader Evaluations

	Effectiveness		Typicality	
Intercept	3.486***	1.893***	3.362***	2.444***
	(0.080)	(0.412)	(0.070)	(0.695)
Leader is a Woman	0.127	0.238*	0.242*	0.328***
	(0.111)	(0.096)	(0.100)	(0.094)
Leader is Gay	0.126	0.069	0.146	0.107
	(0.111)	(0.110)	(0.099)	(0.095)
Woman Leader $\times$ Gay Leader	0.086	-0.015	-0.216	-0.293*
	(0.156)	(0.138)	(0.147)	(0.139)
Controls	No	Yes	No	Yes
N	480	454	490	458

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

The interaction between leader gender and sexuality provides some suggestive evidence of an interactive effect. While there is little evidence of an interaction effect for perceived effectiveness, the estimates for leader typicality indicate that gay women leaders may be viewed as less typical than would be implied by the separate effects of gender and sexuality. Although these estimates are only marginally significant, they point to the possibility that combined identities shape perceptions in ways that are not captured by additive effects alone.

Restricting the sample to participants who pass the manipulation check yields a similar overall pattern, though some point estimates become more pronounced. As in the full sample, there are no statistically significant relationships between leader gender or sexual orientation and follower performance outcomes across measures of item completion and accuracy (Tables 11 and 12). However, the direction of several coefficients shifts. In contrast to the baseline results, where gay leaders were weakly associated with lower output, the manipulation-check sample shows consistently positive coefficients across most performance

measures. For example, gay leaders are associated with higher levels of strict item completion and, in one specification, a statistically significant increase of roughly 4 percentage points in strict accuracy. Although these effects attenuate with controls and remain imprecise overall, they suggest that participants who correctly perceive the treatment may respond more favorably to gay leaders in terms of performance.

Results for leader evaluations remain broadly consistent with the baseline findings but are somewhat stronger in magnitude (Table 13) in this sample. Female leaders are again rated as more effective and more typical, with larger and more precisely estimated coefficients than in the full sample. There is also some suggestive evidence that gay leaders are viewed as more typical in the uncontrolled specification, though this effect does not persist with controls. The interaction between gender and sexuality remains negative for typicality, indicating that gay women leaders may be perceived as less typical relative to the additive effects of gender and sexuality. Overall, conditioning on passing the manipulation check sharpens some patterns, particularly the positive association between gay leaders and performance and the stronger evaluation effects for women, but the main conclusion remains unchanged: leader identity has limited and inconsistent effects on follower performance, even though it may meaningfully shape perceptions.

5.2 Page-Level Performance Differences By Leader Identity

The participant-level results above aggregate performance across all pages completed by each individual. Since all participants were given the same order of book publication pages, this approach does not account for the fact that task difficulty may vary across pages. As a result, differences in participant accuracy may partly reflect exposure to different segments of the task rather than differences in performance conditional on the same task. Participants who persist longer are more likely to encounter later pages, which may not be identical in realized difficulty. This introduces a potential confound between treatment effects and order effects that cannot be addressed in the participant-level analysis.

To address this concern, Table ?? presents page-level estimates that include page fixed effects and cluster standard errors at the participant level. This approach compares participants across treatment conditions within the same page, effectively holding constant the difficulty of the task being performed. In this setting, the outcomes capture within-task performance, such as accuracy and speed on a given page, rather than overall persistence or total output across the experiment.

Table 6: Effects of Leader Characteristics on Page-Level Performance

	Output		Accuracy		Speed		Completed	
Leader is a Woman	-0.019 (0.081)	0.000 (0.085)	-0.010 (0.015)	-0.009 (0.017)	-0.046 (0.082)	-0.020 (0.083)	0.051 (0.099)	0.079 (0.096)
Leader is Gay	-0.059 (0.079)	-0.029 (0.073)	-0.018 (0.016)	-0.013 (0.014)	-0.061 (0.081)	-0.033 (0.077)	0.073 (0.079)	0.042 (0.071)
Woman Leader $\times$ Gay Leader	0.020 (0.121)	-0.038 (0.121)	0.013 (0.024)	0.000 (0.025)	0.050 (0.121)	0.005 (0.123)	-0.096 (0.135)	-0.064 (0.131)
Page FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE (Participant)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Participant Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	4,741	4,163	4,515	3,977	4,735	4,158	4,741	4,163

Standard errors clustered at the participant level in parentheses

All models include page fixed effects

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Across all specifications, the page-level results mirror the participant-level findings. There are no statistically significant relationships between leader gender or sexual orientation and any of the performance outcomes, including output, accuracy, speed, or completion. The estimated coefficients are small in magnitude and unstable in sign across specifications. These findings reinforce the conclusion that leader identity has limited and inconsistent effects on follower behavior, both in aggregate performance and in within-task execution.

5.3 Heterogeneity By Participant Identity and Beliefs

To explore potential mechanisms, I re-estimate the main specifications separately by participant gender and sexuality. I focus my discussion on leader evaluations rather than performance outcomes given that there is little evidence that the null effects of identity on performance mask interesting heterogeneity (see Appendix Tables 11,12, 16, and 17 for the performance heterogeneity results). Across subgroups, the most consistent patterns show

up in leader effectiveness and typicality rather than in any of the performance measures. When I split by participant gender, female leaders tend to be rated more favorably, especially by female participants. Among women, female leaders receive higher ratings on both effectiveness and typicality, with the latter showing clearer statistical support. In contrast, male participants do not show much of a systematic response to leader gender on either dimension. This suggests that the positive evaluation effects in the main results are largely being driven by women’s responses rather than reflecting a general preference across the full sample.

Table 7: Effects of Leader Gender and Sexuality on Leader Evaluations by Participant Gender

	Effectiveness		Typicality	
	Men Only	Women Only	Men Only	Women Only
Leader is a Woman	0.090 (0.179)	0.237 (0.138)	0.202 (0.202)	0.421** (0.131)
Leader is Gay	-0.030 (0.333)	0.038 (0.166)	-0.231 (0.227)	0.206 (0.145)
Woman Leader $\times$ Gay Leader	0.064 (0.410)	0.088 (0.202)	0.224 (0.353)	-0.456* (0.221)
Controls	Yes	Yes	Yes	Yes
N	97	207	97	211

Heteroskedasticity-robust standard errors in parentheses
Control coefficients omitted for brevity
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

A similar pattern emerges when splitting by participant sexuality. Among straight participants, female leaders are again rated as more effective and more typical, lining up with the main results. Among gay participants, the estimates are noisier and less consistent, although there is some suggestive evidence that gay leaders may be viewed as more typical. The interaction between leader gender and sexuality tends to pull typicality ratings down, particularly among women, which suggests that gay women leaders may be seen as less prototypical in the nonprofit setting.

I also investigate these effects by the underlying heteronormative beliefs of participants. Low scores of heteronormativity would indicate beliefs of gender egalitarianism and positive views of non-straight sexualities. I find in Table 9 under a median split of the distribution of participant beliefs that those that fall in the low category drive the higher typicality rating of

Table 8: Effects of Leader Gender and Sexuality on Leader Evaluations by Participant Sexuality

	Effectiveness		Typicality	
	Straight Only	Gay Only	Straight Only	Gay Only
Leader is a Woman	0.286*	0.023	0.290*	0.515.
	(0.122)	(0.239)	(0.122)	(0.268)
Leader is Gay	-0.010	0.207	0.022	0.403
	(0.180)	(0.309)	(0.142)	(0.284)
Woman Leader $\times$ Gay Leader	-0.046	0.279	-0.221	-0.404
	(0.241)	(0.305)	(0.239)	(0.349)
Controls	Yes	Yes	Yes	Yes
N	217	87	218	90

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

women leaders that does not extend to gay women leaders. While the results are noisy, there is evidence in Appendix Tables 18 and 19 that followers with high level of heteronormative beliefs perform more slowly and less accurately for gay male leaders. However, this sample skews toward low levels of heteronormative belief and as such these findings should be treated as suggestive pending replication in more ideologically diverse samples.

Table 9: Effects of Leader Gender and Sexuality on Leader Evaluations by Heteronormativity

	Effectiveness		Typicality	
	Low Hetnorm	High Hetnorm	Low Hetnorm	High Hetnorm
Leader is a Woman	0.172	0.236.	0.459**	0.226
	(0.139)	(0.142)	(0.141)	(0.141)
Leader is Gay	0.112	0.127	0.203	-0.024
	(0.152)	(0.143)	(0.145)	(0.141)
Woman Leader $\times$ Gay Leader	0.013	0.040	-0.526**	-0.012
	(0.195)	(0.208)	(0.200)	(0.212)
Controls	Yes	Yes	Yes	Yes
N	211	212	217	210

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

6 Discussion

A central question of this study is whether leader identity shapes how followers allocate effort and perform on assigned tasks. The evidence provides little support for this relationship. Across specifications, differences in both task completion and accuracy by leader gender are small and statistically indistinguishable from zero. Likewise, leader sexuality has no

meaningful effect on follower performance, with coefficients that are both economically small and statistically insignificant. Evaluations of leaders, however, reveal differences in how leaders are perceived based on their identity. In particular, women leaders tend to receive higher ratings in effectiveness and typicality than male leaders, with this pattern driven primarily by female participants. Strikingly, this boost in ratings does not extend to gay women.

These results highlight a disconnect between stated beliefs and observed behavior. While leader identity has little measurable impact on follower performance, it does influence how leaders are evaluated. This divergence suggests that perceptions of leadership may not translate into meaningful differences in task execution, even when participants express systematic differences in how they view leaders across identity dimensions.

In many real-world organizations, promotion and advancement decisions rely heavily on written evaluations, peer feedback, and subjective assessments of leadership ability. If these stated evaluations are systematically shaped by identity-based perceptions that are only weakly related to actual performance, then they may provide a noisy or even misleading signal of underlying effectiveness. As a result, organizations that rely heavily on such feedback may inadvertently reward perceived leadership qualities rather than realized outcomes. These findings suggest value in incorporating objective performance measures alongside subjective evaluations when making leadership decisions, as well as carefully considering how evaluation systems may amplify or dampen identity-based biases.

An important direction for future research is to directly compare subjective evaluations, such as written feedback or peer ratings, with objective performance metrics in both experimental and field settings. Doing so would help identify when and why these measures diverge, and whether certain evaluation frameworks systematically amplify identity-based perceptions. More importantly, this comparison allows for an assessment of the economic consequences of these differences. If leadership positions are allocated based on subjective evaluations that do not align with true performance, organizations may incur efficiency

losses by misallocating talent. Disentangling these relationships is critical for quantifying the gains or losses associated with how leadership roles are assigned and for designing evaluation systems that better promote efficient outcomes.

Disclosure Statement

The author has no conflicts of interest to declare.

References

- Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity*. *Quarterly Journal of Economics*, 115(3), 715–753. <https://doi.org/10.1162/003355300554881>
- Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19(1), 9–32. <https://doi.org/10.1257/0895330053147930>
- Aksoy, C. G., Carpenter, C. S., Frank, J., & Huffman, M. L. (2019). Gay glass ceilings: Sexual orientation and workplace authority in the UK. *Journal of Economic Behavior & Organization*, 159, 167–180. <https://doi.org/10.1016/j.jebo.2019.01.013>
- Antonakis, J., d’Adda, G., Weber, R. A., & Zehnder, C. (2022). “Just Words? Just Speeches?” On the Economic Value of Charismatic Leadership. *Management Science*, 68(9), 6355–6381. <https://doi.org/10.1287/mnsc.2021.4219>
- Antonakis, J., Fenley, M., & Liechti, S. (2011). Can Charisma Be Taught? Tests of Two Interventions. *Academy of Management Learning & Education*, 10(3), 374–396. <https://doi.org/10.5465/amle.2010.0012>
- Aunga, D., & Masare, O. (2017). Effect of leadership styles on teacher’s performance in primary schools of Arusha District Tanzania. *International Journal of Educational Policy Research and Review*, 4(4).

- Badgett, M. L., Waaldijk, K., & Rodgers, Y. V. D. M. (2019). The relationship between LGBT inclusion and economic development: Macro-level evidence. *World Development*, *120*, 1–14. <https://doi.org/10.1016/j.worlddev.2019.03.011>
- Bass, B. M. (1985). Leadership: Good, better, best. *Organizational Dynamics*, *13*(3), 26–40. [https://doi.org/10.1016/0090-2616\(85\)90028-2](https://doi.org/10.1016/0090-2616(85)90028-2)
- Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87. <https://doi.org/10.1016/j.jebo.2018.02.024>
- Correll, S. J. (2004). Constraints into Preferences: Gender, Status, and Emerging Career Aspirations. *American Sociological Review*, *69*(1), 93–113. <https://doi.org/10.1177/000312240406900106>
- Cotter, D. A., Hermsen, J. M., Ovadia, S., & Vanneman, R. (2001). The Glass Ceiling Effect. *Social Forces*, *80*(2), 655–681. <https://doi.org/10.1353/sof.2001.0091>
- Eagly, A. H., & Chin, J. L. (2010). Diversity and leadership in a changing world. *American Psychologist*, *65*(3), 216–224. <https://doi.org/10.1037/a0018957>
- Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, *129*(4), 569–591. <https://doi.org/10.1037/0033-2909.129.4.569>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Fasoli, F., Maass, A., Paladino, M. P., & Sulpizio, S. (2017). Gay- and Lesbian-Sounding Auditory Cues Elicit Stereotyping and Discrimination. *Archives of Sexual Behavior*, *46*(5), 1261–1277. <https://doi.org/10.1007/s10508-017-0962-0>
- Fest, S., Kvaløy, O., Nieken, P., & Schöttner, A. (2021). How (not) to motivate online workers: Two controlled field experiments on leadership in the gig economy. *The Leadership Quarterly*, *32*(6), 101514. <https://doi.org/10.1016/j.leaqua.2021.101514>

- Fischer, T., Dietz, J., & Antonakis, J. (2017). Leadership Process Models: A Review and Synthesis. *Journal of Management*, 43(6), 1726–1753. <https://doi.org/10.1177/0149206316682830>
- Frank, J. (2006). Gay Glass Ceilings. *Economica*, 73(291), 485–508. <https://doi.org/10.1111/j.1468-0335.2006.00516.x>
- Gardner, W. L., Coglisier, C. C., Davis, K. M., & Dickens, M. P. (2011). Authentic leadership: A review of the literature and research agenda. *The Leadership Quarterly*, 22(6), 1120–1145. <https://doi.org/10.1016/j.leaqua.2011.09.007>
- Giacomin, M., Tskhay, K. O., & Rule, N. O. (2022). Gender stereotypes explain different mental prototypes of male and female leaders. *The Leadership Quarterly*, 33(6), 101578. <https://doi.org/10.1016/j.leaqua.2021.101578>
- Greenleaf, R. K., & Spears, L. C. (1998). *The power of servant-leadership: Essays*. Berrett-Koehler Publishers.
- Grossman, S. J., & Hart, O. D. (1992). An Analysis of the Principal-Agent Problem [Series Title: Huebner International Series on Risk, Insurance and Economic Security]. In *Foundations of Insurance Economics* (pp. 302–340, Vol. 14). Springer Netherlands. https://doi.org/10.1007/978-94-015-7957-5_16
- Habarth, J. M. (2015). Development of the heteronormative attitudes and beliefs scale. *Psychology & Sexuality*, 6(2), 166–188. <https://doi.org/10.1080/19419899.2013.876444>
- Hanel, P. H. P., & Vione, K. C. (2016). Do Student Samples Provide an Accurate Estimate of the General Public? (M. Voracek, Ed.). *PLOS ONE*, 11(12), e0168354. <https://doi.org/10.1371/journal.pone.0168354>
- Heilman, M. E. (2001). Description and Prescription: How Gender Stereotypes Prevent Women’s Ascent Up the Organizational Ladder. *Journal of Social Issues*, 57(4), 657–674. <https://doi.org/10.1111/0022-4537.00234>

- Hooper, A. (Ed.). (2016). *Leadership perspectives* [OCLC: 988029023]. Routledge, Taylor & Francis Group.
- Kim, S. (2002). Participative Management and Job Satisfaction: Lessons for Management Leadership. *Public Administration Review*, 62(2), 231–241. <https://doi.org/10.1111/0033-3352.00173>
- Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, 137(4), 616–642. <https://doi.org/10.1037/a0023557>
- Lazear, E. P. (2012). Leadership: A personnel economics approach. *Labour Economics*, 19(1), 92–101. <https://doi.org/10.1016/j.labeco.2011.08.005>
- Liberman, B. E., & Golom, F. D. (2015). Think manager, think male? Heterosexuals’ stereotypes of gay and lesbian managers. *Equality, Diversity and Inclusion: An International Journal*, 34(7), 566–578. <https://doi.org/10.1108/EDI-01-2015-0005>
- Mallory, C., Sears, B., & Moran, R. (2024, August). *LGBTQ People’s Experiences of Workplace Discrimination and Harassment* (tech. rep.). The Williams Institute, UCLA School of Law. <https://williamsinstitute.law.ucla.edu/wp-content/uploads/Workplace-Discrimination-Aug-2024.pdf>
- McManus, R. M., & Perruci, G. (Eds.). (2020). *Understanding leadership: An arts and humanities perspective* (Second Edition). Routledge.
- Revised edition of Understanding leadership, 2015.
- Meslec, N., Curseu, P. L., Fodor, O. C., & Kenda, R. (2020). Effects of charismatic leadership and rewards on individual performance. *The Leadership Quarterly*, 31(6), 101423. <https://doi.org/10.1016/j.leaqua.2020.101423>
- Morton, J. W. (2017). Think leader, think heterosexual male? The perceived leadership effectiveness of gay male leaders. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l’Administration*, 34(2), 159–169. <https://doi.org/10.1002/cjas.1434>

- Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*, 33(1), 31–88. <https://doi.org/10.1145/375360.375365>
- Niedlich, C., & Steffens, M. C. (2015). On the interplay of (positive) stereotypes and prejudice: Impressions of lesbian and gay applicants for leadership positions. *Sensoria: A Journal of Mind, Brain & Culture*, 11(1). <https://doi.org/10.7790/sa.v11i1.408>
- Paustian-Underdahl, S. C., Walker, L. S., & Woehr, D. J. (2014). Gender and perceptions of leadership effectiveness: A meta-analysis of contextual moderators. *Journal of Applied Psychology*, 99(6), 1129–1145. <https://doi.org/10.1037/a0036751>
- Pellegrini, V., De Cristofaro, V., Giacomantonio, M., & Salvati, M. (2020). Why are gay leaders perceived as ineffective? The role of the type of organization, sexual prejudice and gender stereotypes. *Personality and Individual Differences*, 157, 109817. <https://doi.org/10.1016/j.paid.2020.109817>
- Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and Managerial Stereotypes: Have the Times Changed? *Journal of Management*, 28(2), 177–193. <https://doi.org/10.1177/014920630202800203>
- Puni, A., Ofei, S. B., & Okoe, A. (2014). The Effect of Leadership Styles on Firm Performance in Ghana. *International Journal of Marketing Studies*, 6(1), p177. <https://doi.org/10.5539/ijms.v6n1p177>
- Ridgeway, C. L. (2001). Gender, Status, and Leadership. *Journal of Social Issues*, 57(4), 637–655. <https://doi.org/10.1111/0022-4537.00233>
- Rosette, A. S., & Livingston, R. W. (2012). Failure is not an option for Black women: Effects of organizational performance on leaders with single versus dual-subordinate identities. *Journal of Experimental Social Psychology*, 48(5), 1162–1167. <https://doi.org/10.1016/j.jesp.2012.05.002>
- Ryan, M. K., & Haslam, S. A. (2005). The Glass Cliff: Evidence that Women are Over-Represented in Precarious Leadership Positions. *British Journal of Management*, 16(2), 81–90. <https://doi.org/10.1111/j.1467-8551.2005.00433.x>

- Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016). Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly*, 27(3), 446–455. <https://doi.org/10.1016/j.leaqua.2015.10.008>
- Schein, V. E. (2001). A Global Look at Psychological Barriers to Women’s Progress in Management. *Journal of Social Issues*, 57(4), 675–688. <https://doi.org/10.1111/0022-4537.00235>
- Sheer, V. C. (2015). “Exchange lost” in leader–member exchange theory and research: A critique and a reconceptualization. *Leadership*, 11(2), 213–229. <https://doi.org/10.1177/1742715014530935>
- Snyder, K. (2006). *The G quotient: Why gay executives are excelling as leaders– and what every manager needs to know* (1st ed) [OCLC: ocm65340952]. Jossey-Bass.
- Wild, H., Kyröläinen, A.-J., & Kuperman, V. (2022). How representative are student convenience samples? A study of literacy and numeracy skills in 32 countries (S. Frisson, Ed.). *PLOS ONE*, 17(7), e0271191. <https://doi.org/10.1371/journal.pone.0271191>
- Winkler, W. (1990). *String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage*. (tech. rep.). US Census.

Appendix

A Training Time

This appendix presents estimates of the effect of leader identity on the time participants spent on the training materials. Across specifications, neither leader gender nor sexual orientation is associated with statistically significant differences in training time. The coefficient on female leader is negative and moderately sized but imprecisely estimated, and the gay leader coefficient is near zero. These results suggest that leader identity does not influence on average how much effort participants invest in preparing for the task, consistent with the

null performance results in the main text.

Table 10: Effects of Leader Gender and Sexuality on Training Time

	Training Time
Intercept	-62.317 (54.808)
Leader is a Woman	-22.214 (20.604)
Leader is Gay	2.291 (14.613)
Woman Leader $\times$ Gay Leader	9.485 (18.879)
Controls	Yes
N	454
Robust standard errors in parentheses	
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$	

B Sample Subset: Passed Manipulation Check

Tables 11 through 13 replicate the main analyses restricting the sample to participants who correctly identified their assigned leader’s gender and sexual orientation.

For item completion and accuracy (Tables 11 and 12), the pattern of null results from the main analysis is broadly preserved. The coefficient on gay leader becomes positive across most performance specifications. This pattern is in the opposite direction from what a discrimination-based account would predict, and may reflect that participants who attended carefully to their leader’s identity responded more favorably to gay leaders in terms of task effort. These estimates are imprecise, however, and should be interpreted cautiously.

Leader evaluation results (Table 13) are larger in magnitude in this subsample than in the full sample. Female leaders are rated as more effective and more typical with larger and more precisely estimated coefficients, consistent with the main results. The interaction between gay and female leader remains negative for typicality, reinforcing the finding that gay women leaders are perceived as less prototypical than would be implied by the additive effects of gender and sexuality alone.

Table 11: Effects of Leader Gender and Sexuality on Item Completion

	Items		Max Items		Strict Items	
Intercept	23.658***	22.163***	32.342***	31.070***	21.423***	19.255**
	(0.668)	(6.070)	(0.971)	(7.785)	(0.706)	(6.119)
Leader is a Woman	0.117	-0.243	-0.090	-0.448	0.135	-0.145
	(1.025)	(0.955)	(1.488)	(1.333)	(1.022)	(0.992)
Leader is Gay	0.417	0.659	0.389	1.050	1.293	1.395
	(1.058)	(1.073)	(1.535)	(1.527)	(1.094)	(1.102)
Woman Leader $\times$ Gay Leader	0.975	0.551	1.071	0.278	-0.018	-0.517
	(1.635)	(1.509)	(2.346)	(2.138)	(1.628)	(1.544)
Controls	No	Yes	No	Yes	No	Yes
Passed Manipulation Check	Yes	Yes	Yes	Yes	Yes	Yes
N	355	330	355	330	355	330

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

Table 12: Effects of Leader Gender and Sexuality on Accuracy

	Pct. Correct		Max Pct.		Strict Pct.	
Intercept	59.949***	34.813**	80.935***	49.410**	54.209***	27.932*
	(1.005)	(11.464)	(1.135)	(15.811)	(1.308)	(12.422)
Leader is a Woman	0.418	-0.088	-0.502	-0.980	1.081	0.685
	(1.539)	(1.610)	(1.760)	(1.766)	(1.898)	(2.009)
Leader is Gay	1.875	-0.126	2.266	0.147	4.002*	1.535
	(1.568)	(1.645)	(1.747)	(1.868)	(1.940)	(2.027)
Woman Leader $\times$ Gay Leader	1.061	1.956	0.856	1.445	-1.578	-1.008
	(2.346)	(2.421)	(2.463)	(2.509)	(2.873)	(2.964)
Controls	No	Yes	No	Yes	No	Yes
Passed Manipulation Check	Yes	Yes	Yes	Yes	Yes	Yes
N	355	330	355	330	355	330

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

Table 13: Effects of Leader Gender and Sexuality on Leader Evaluations

	Effectiveness		Typicality	
Intercept	3.502***	1.522***	3.383***	2.287**
	(0.086)	(0.400)	(0.074)	(0.860)
Leader is a Woman	0.210.	0.268*	0.299**	0.342***
	(0.117)	(0.104)	(0.106)	(0.103)
Leader is Gay	0.223	0.043	0.252*	0.095
	(0.142)	(0.145)	(0.121)	(0.113)
Woman Leader $\times$ Gay Leader	-0.025	-0.011	-0.325.	-0.332.
	(0.195)	(0.176)	(0.185)	(0.175)
Controls	No	Yes	No	Yes
Passed Manipulation Check	Yes	Yes	Yes	Yes
N	337	326	341	330

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

C Sample Subset: Participant Gender

Table 14: Effects of Leader Gender and Sexuality on Item Completion by Participant Gender

	Items		Max Items		Strict Items	
	Men Only	Women Only	Men Only	Women Only	Men Only	Women Only
Leader is a Woman	2.012 (1.665)	-0.979 (1.173)	2.929 (2.399)	-1.954 (1.605)	2.538 (1.695)	-0.948 (1.246)
Leader is Gay	-0.010 (1.954)	1.369 (1.334)	-1.312 (2.497)	1.954 (1.935)	1.880 (1.973)	1.987 (1.366)
Woman Leader $\times$ Gay Leader	-1.061 (2.678)	0.622 (1.853)	-1.378 (3.593)	0.480 (2.559)	-3.542 (2.801)	-0.007 (1.894)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	97	212	97	212	97	212

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

Table 15: Effects of Leader Gender and Sexuality on Item Completion by Participant Gender

	Items		Max Items		Strict Items	
	Men Only	Women Only	Men Only	Women Only	Men Only	Women Only
Leader is a Woman	2.012 (1.665)	-0.979 (1.173)	2.929 (2.399)	-1.954 (1.605)	2.538 (1.695)	-0.948 (1.246)
Leader is Gay	-0.010 (1.954)	1.369 (1.334)	-1.312 (2.497)	1.954 (1.935)	1.880 (1.973)	1.987 (1.366)
Woman Leader $\times$ Gay Leader	-1.061 (2.678)	0.622 (1.853)	-1.378 (3.593)	0.480 (2.559)	-3.542 (2.801)	-0.007 (1.894)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	97	212	97	212	97	212

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

D Sample Subset: Participant Sexuality

Table 16: Effects of Leader Gender and Sexuality on Item Completion by Participant Sexuality

	Items		Max Items		Strict Items	
	Straight Only	Gay Only	Straight Only	Gay Only	Straight Only	Gay Only
Leader is a Woman	-0.074 (1.119)	0.444 (2.067)	-0.556 (1.534)	0.914 (2.985)	0.580 (1.139)	0.301 (2.260)
Leader is Gay	0.825 (1.194)	1.338 (2.624)	1.218 (1.605)	1.309 (4.024)	1.344 (1.176)	3.488 (2.774)
Woman Leader $\times$ Gay Leader	1.135 (1.745)	-2.500 (3.235)	0.519 (2.367)	-2.980 (4.785)	-0.177 (1.806)	-4.133 (3.320)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	219	90	219	90	219	90

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

Table 17: Effects of Leader Gender and Sexuality on Accuracy by Participant Sexuality

	Pct. Correct		Max Pct.		Strict Pct.	
	Straight Only	Gay Only	Straight Only	Gay Only	Straight Only	Gay Only
Leader is a Woman	-0.008 (1.944)	0.464 (2.846)	-1.603 (2.124)	1.526 (3.467)	2.062 (2.333)	0.540 (4.098)
Leader is Gay	-0.570 (1.955)	2.757 (3.743)	-0.340 (2.349)	2.872 (4.332)	0.652 (2.200)	7.767 (4.867)
Woman Leader $\times$ Gay Leader	3.680 (2.888)	-1.743 (4.291)	2.609 (3.223)	-2.495 (4.691)	-0.271 (3.506)	-5.762 (5.884)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	219	90	219	90	219	90

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

E Sample Subset: Low vs High Heteronormative Beliefs

Table 18: Effects of Leader Gender and Sexuality on Item Completion by Heteronormativity

	Items		Max Items		Strict Items	
	Low Hetnorm	High Hetnorm	Low Hetnorm	High Hetnorm	Low Hetnorm	High Hetnorm
Leader is a Woman	0.850 (1.277)	-0.048 (1.382)	0.478 (1.780)	-0.045 (1.879)	0.639 (1.352)	0.666 (1.426)
Leader is Gay	-0.209 (1.379)	-1.586 (1.420)	-0.176 (1.836)	-1.021 (1.989)	0.571 (1.513)	-1.729 (1.458)
Woman Leader $\times$ Gay Leader	-1.185 (1.873)	0.114 (2.071)	-1.138 (2.600)	-0.877 (2.819)	-1.872 (1.980)	-0.684 (2.120)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	217	212	217	212	217	212

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$

Table 19: Effects of Leader Gender and Sexuality on Accuracy by Heteronormativity

	Pct. Correct		Max Pct.		Strict Pct.	
	Low Hetnorm	High Hetnorm	Low Hetnorm	High Hetnorm	Low Hetnorm	High Hetnorm
Leader is a Woman	1.606 (1.857)	0.496 (2.271)	0.835 (1.889)	-0.259 (2.578)	1.472 (2.601)	2.538 (2.678)
Leader is Gay	-1.743 (2.121)	-2.251 (2.038)	-1.324 (2.226)	-0.632 (2.475)	-0.047 (2.698)	-2.295 (2.564)
Woman Leader $\times$ Gay Leader	-1.719 (2.805)	0.634 (3.179)	-2.228 (2.885)	-0.270 (3.679)	-3.732 (3.703)	-1.951 (3.923)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	217	212	217	212	217	212

Heteroskedasticity-robust standard errors in parentheses

Control coefficients omitted for brevity

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, . $p < 0.15$