



A framing effect in the judgment of discrimination

Christopher K. Hsee^{a,1} and Xilin Li^{b,1}

Edited by Susan Fiske, Princeton University, Princeton, NJ; received April 12, 2022; accepted September 11, 2022

Discrimination is not only an objective fact but also a subjective judgment. While extensive research has studied discrimination as an objective fact, we study the judgment of discrimination and show that it is malleable while holding objective discrimination constant. We focus on a common situation in real life: the constituent groups in a candidate pool are unequal (e.g., fewer female candidates than male candidates for tech jobs), and observers (e.g., the public) see only one side of the decision outcome (e.g., only the hired applicants, not the rejected ones). Ten experiments reveal a framing effect: people judge the decision-maker (e.g., the tech firm) as more discriminatory against the minority in the candidate pool if people see the composition of the *accepted* candidates than if they see the composition of the rejected candidates, even though the information in the two frames is equivalent (i.e., knowing the information in one frame is sufficient to infer the information in the other). The framing effect occurs regardless of whether the decision-maker is objectively discriminatory, replicates across diverse samples (Americans, Asians, and Europeans) and types of discrimination (e.g., gender, race, political orientation), and has significant behavioral consequences. We theorize and show that the framing effect arises because, when judging discrimination, people overlook information that they could infer but is not explicitly given, and they expect equality in the composition of the constituent groups in their given frame. This research highlights the fallibility of judged discrimination and suggests interventions to reduce biases and increase accuracy.

gender discrimination | racial discrimination | judgment biases | accept-reject framing | hiring decisions

Discrimination is prevalent and serious. To fight discrimination, people need to judge discrimination accurately, however, judgments may be biased. While extensive existing research has studied and demonstrated real discrimination [e.g., (1–5)], we study the judgment of discrimination. (We use the term discrimination broadly to include prejudice and the opposite of favoritism.) Understanding the judgment of discrimination is valuable because it is often people's subjective judgments (i.e., their perceptions) that form public opinions, affect psychological well-being, and health (6, 7), and lead to social and political changes.

In particular, we study the judgment of discrimination in situations where the constituent groups in a candidate pool are not equally represented (e.g., fewer women than men), and people (observers) see only one side of the decision outcome—either the composition of the accepted candidates or the composition of the rejected candidates, not both.

These situations are common in real life. For example, there are usually fewer women than men among applicants for science, technology, engineering, and mathematics (STEM jobs, as fewer women than men have degrees in these fields (8, 9). Also, the public usually sees only the composition of the accepted candidates (how many men and women were hired) (10), not the composition of the rejected candidates (how many men and women were rejected). (We simply assume that the composition of the candidate pool is unequal, and we do not investigate why. The causes of unbalanced candidate pools are important but beyond the scope of this research.)

In such situations, we propose and show that the decision-maker will be judged as more discriminatory against the minority in the candidate pool (and less discriminatory against the majority in the candidate pool) if people (e.g., the public) see the composition of the accepted candidates rather than the composition of the rejected candidates. This framing effect arises even though the information in the two frames is equivalent (i.e., knowing the information in one frame is sufficient to infer the information in the other).

To clarify, the purpose of this article is not to downplay the prevalence and seriousness of real discrimination. Rather, we show that judged discrimination is malleable while objective discrimination is held constant.

Paradigm

To investigate the judgment of discrimination scientifically, all our experiments adopt the following experimental paradigm (or a variant). We randomly assign research

Significance

Discrimination is prevalent and serious. To detect and fight discrimination while avoiding making false accusations, people need to judge discrimination accurately, however, judgments are not always accurate. We document a robust bias in judged discrimination—people deem a decision-maker (e.g., a firm making hiring decisions) as more discriminatory against the minority in a candidate pool (e.g., female candidates) if they see the composition of the accepted candidates (e.g., how many men and women were hired) than if they see the composition of the rejected candidates. This effect occurs regardless of whether or not the decision-maker is objectively discriminatory. We explain why this bias occurs and discuss how to mitigate the bias and increase accuracy in the judgment of discrimination.

Author affiliations: ^aUniversity of Chicago Booth School of Business, Chicago, IL 60637; and ^bChina Europe International Business School (CEIBS), Shanghai, China

Author contributions: Both authors developed the idea and designed the studies; X.L. collected and analyzed the data; C.K.H. wrote the original draft; both authors revised the article.

The author declares no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: Christopher.Hsee@chicagobooth.edu or xilinli@ceibs.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2205988119/-DCSupplemental>.

Published November 14, 2022.

participants to one of two framing conditions: accept and reject. All participants are told that a certain candidate pool contained two equally qualified but unbalanced groups (e.g., 20% of the candidates were female and 80% were male), and the decision-maker accepted half the candidates and rejected the other half. In addition, participants in the accept condition are informed of the composition of the accepted candidates (e.g., 20% of the accepted candidates were female and 80% were male), whereas participants in the reject condition are informed of the composition of the rejected candidates (e.g., 20% of the rejected candidates were female and 80% were male). Finally, all participants are asked whether the decision-maker exhibited discrimination; participants answer on a continuous bipolar scale ranging from strong discrimination against the minority to strong discrimination against the majority.

For simplicity, we will use the following notations in our analyses (but not in our instructions to participants): Min and Maj refer to the minority and majority groups, respectively, in the candidate pool. (Unless otherwise specified, we use minority and majority to refer to the minority and majority members in the candidate pool.) $Min\%_C$, $Min\%_A$, and $Min\%_R$ describe the composition (i.e., the proportion of Min members) of the candidate pool, the accepted candidates, and the rejected candidates, respectively. (Since the proportions of Maj members are simply 100% minus the proportions of Min members, we do not include notations for Maj members in our analyses. In our experiments, however, we always provided participants with the statistics of both Min and Maj members.)

Using these notations, we summarize our paradigm as follows. All participants are informed of $Min\%_C$ and know that the decision-maker accepted half the candidates and rejected half the candidates. Also, those in the accept condition are informed of $Min\%_A$, and those in the reject condition are informed of $Min\%_R$.

Note that the information in the two conditions is equivalent. Since everyone is informed of $Min\%_C$ and knows that the decision-maker accepted half the candidates and rejected half the candidates, participants in the accept condition (who are informed of $Min\%_A$) can easily infer $Min\%_R$, and vice versa. In most of our studies, we simplify the inference process by setting both $Min\%_A$ and $Min\%_R$ equal to $Min\%_C$, so virtually no calculation is required to make the inference. For instance, if the proportion of women in the candidate pool ($Min\%_C$) is 25%, and the proportion of women in the accepted pool ($Min\%_A$) is 25%, then it is obvious that the proportion of women in the rejected pool ($Min\%_R$) must be 25%, too.

Theory and Hypothesis

Normatively, to judge discrimination, people should consider both the composition of the accepted candidates ($Min\%_A$) and the composition of the rejected candidates ($Min\%_R$). Specifically, they should judge $Min\%_A$ relative to $Min\%_R$: the smaller (the greater) $Min\%_A$ is relative to $Min\%_R$, the more the decision-maker discriminated against Min members (against Maj members).

For ease of exposition, we define a decision-maker as showing no objective discrimination if $Min\%_A = Min\%_R$ (e.g., if the proportion of females among those accepted by the decision-maker equals the proportion of females among those rejected by the decision-maker). We wish to clarify that no objective discrimination may not be the best practice in every situation. In some situations (e.g., when a tech firm with predominantly male employees wants to increase the representation of women in the

workforce), the best practice may be to have some objective discrimination (e.g., such that females comprise a larger proportion of the accepted pool than of the rejected pool, i.e., $Min\%_A > Min\%_R$). Importantly, however, regardless of what the best practice is, the normative judgment of discrimination should be based on both $Min\%_A$ and $Min\%_R$, so it should not be swayed by framing.

However, we predict that actual judgments of discrimination are significantly influenced by framing. We theorize that, when judging discrimination, people do not compare $Min\%_A$ with $Min\%_R$ (as they should). Instead, they compare the composition in their condition with some “expected” (i.e., ideal) composition, and they usually (but not always) expect equal representation of the constituent groups (e.g., equal proportions of men and women).

Specifically, we propose that when judging discrimination, people have two psychological tendencies: flipside neglect and the equality expectation. By flipside neglect, we mean that people focus on the information given in their condition ($Min\%_A$ in the accept condition; $Min\%_R$ in the reject condition) and overlook the information in the other condition, even though people could easily infer it. This proposition is based on existing research showing that decision-makers often overlook information that is important but not explicitly given, even though they can infer the information from what they know. This general tendency is what Kahneman refers to as “what you see is all there is” (11).*

By the equality expectation, we mean that people use equality (50%) as a default reference point with which they judge the composition in their condition. This proposition builds on existing research showing that equality is often the default benchmark for judging whether someone is fair or prejudiced (27–29). If the observed composition deviates from the equality expectation, people will likely consider the decision-maker discriminatory. The direction of the judgment depends on the condition. If you are in the accept condition and find out that 80% of the candidates accepted by a company are men, and only 20% are women, you will likely consider the company discriminatory against women because most of the accepted candidates are men. If you are in the reject condition and find out that 80% of the candidates rejected by the company are men, and only 20% are women, you will likely consider the company discriminatory against men because most of the rejected candidates are men. In other words, your judgment of the company depends on whether you are in the accept or reject condition. (Equality is only a default expectation; people do not always expect equality. We will discuss other expectations later.)

Formally, flipside neglect means that judged discrimination in the accept condition depends on $Min\%_A$ (not $Min\%_R$), and judged discrimination in the reject condition depends on $Min\%_R$ (not $Min\%_A$). The equality expectation means that people use 50% as the default reference point for judging discrimination, in both conditions. Thus, the formal models are:

$$JD_A = -(Min\%_A - 50\%) + k \quad [1A]$$

$$JD_R = (Min\%_R - 50\%) + k \quad [1R]$$

*At this general level, our framing effect resembles many existing findings, including (but not limited to) one-sided evaluation (12), illusory correlation (13, 14), base rate neglect (15–17), selection bias (18–20), narrow bracketing (21–24), and the choose-reject framing effect (25, 26). Upon closer inspection, however, these findings are different from each other and from ours. For example, one-sided evaluation typically occurs in situations where information on the other side is unattainable, but our effect occurs even when the information on the other side can be easily inferred. The illusory correlation is a discrepancy between a judged correlation and the actual correlation (e.g., people overestimate the extent to which infrequent undesirable behaviors are performed by minorities rather than by the majority), whereas our effect is about framing—a discrepancy in judged discrimination between two equivalent frames. The choose-reject framing effect (25, 26) arises from choice conflicts, but our effect does not involve choice conflicts.

where JD_A and JD_R are judged discrimination in the accept and reject conditions, respectively. We define both variables such that greater values represent stronger judged discrimination against Min members (or weaker judged discrimination against Maj members). The two equations have opposite signs because, by definition, JD_A is negatively correlated with $Min\%_A$, while JD_R is positively correlated with $Min\%_R$. (We include k at the end to capture factors unrelated to our theory. Although k may influence the absolute levels of JD_A and JD_R , it does not influence the phenomenon of interest—the framing effect, namely, the difference between JD_A and JD_R .)

Given Eqs. **1A** and **1R**, the difference between JD_A and JD_R is:

$$JD_A - JD_R = 100\% - (Min\%_A + Min\%_R) \quad [2]$$

For ease of computation, we assume in our analysis (and we told participants in our experiments) that the decision-maker accepted half the candidates and rejected half the candidates. (Even if the accept and reject rates are not half the candidate pool, the analysis still holds as long as there is no objective discrimination.) This means *that on average*, the composition of the accepted candidates and the composition of the rejected candidates are the same as the composition of the candidate pool, namely:

$$(Min\%_A + Min\%_R)/2 = Min\%_C \quad [3]$$

Combining Eqs. **2** and **3** yields:

$$JD_A - JD_R = 100\% - 2 \times Min\%_C \quad [4]$$

which means:

$$JD_A > JD_R \text{ if } Min\%_C < 50\% \quad \text{Inequality 1}$$

Inequality 1 is our central hypothesis. It predicts that if the candidate pool contains unequal constituent groups (i.e., $Min\%_C < 50\%$), then people in the accept condition will judge the decision-maker as more discriminatory against Min members (and less against Maj members) than people in the reject condition. This is our proposed framing effect. The framing effect will occur even if the decision-maker is objectively discriminatory and, if so, regardless of whether the decision-maker objectively discriminates against Min or Maj members.[†]

Notably, our theory predicts only that $JD_A > JD_R$. It does not comment on the absolute levels of JD_A or JD_R (e.g., does not predict whether $JD_A > 0$ or $JD_R < 0$) because our theory is agnostic about the k factor in Eqs. **1A** and **1R**. In other words, our theory predicts only a framing effect and is silent about whether the judgment in either framing condition alone is biased. Even so, we can conclude that if people render different judgments in the two (normatively equivalent) framing conditions, their judgment in at least one of the conditions must be biased. The presence of a framing effect is evidence that judged discrimination is fallible.

Materials, Methods, and Results

We tested our proposed framing effect, its behavioral consequences, and its moderators in ten studies, with diverse samples

[†]Our paradigm and findings differ from those of other researchers who studied accept-versus-reject effects in discrimination-related domains. For example, Hugenberg et al. (30) find that asking a decision-maker to adopt different strategies (accept vs. reject) influences the decision-maker's own attitudes, whereas we find that using different frames to describe someone else's decision influences the observer's judgment. Phillips and Jun (31) find that an objectively discriminatory decision is considered less discriminatory when framed as accepting advantaged candidates (e.g., accepting majority members) than as rejecting disadvantaged candidates (e.g., rejecting minority members), whereas we find that any decision—regardless of its objective discrimination—is considered more discriminatory when framed as an acceptance decision for both minority and majority members than as a rejection decision for both minority and majority members.

and mixed methods. See Table 1 for an overview, including the key findings. Because most studies followed similar procedures, we describe the studies only briefly below. For detailed descriptions, see *SI Appendix, SI Part 1*; for a single-paper meta-analysis of the key findings, see *SI Appendix, SI Part 2*. All studies were approved by the IRB of the University of Chicago, and all participants provided informed consent.

Study 1A demonstrated the framing effect using a nationally representative sample from the United States ($n = 448$; recruited on Prime Panels using a US census-compatibility sample). Participants were randomly assigned to the accept and reject conditions. All participants read a case in which a private school recently recruited new teachers. The candidate pool was 20% female and 80% male; the constituent groups were equally qualified, and the principal of the school knew that. The principal hired half the applicants and rejected half the applicants. Also, participants in the [accept] {reject} condition were told that 20% of the applicants [hired] {rejected} by the principal were female, and 80% were male. All participants were asked to evaluate the principal's discrimination.

Note that female applicants comprised the same proportions of the accepted and rejected pools (i.e., $Min\%_A = Min\%_R = Min\%_C = 20\%$), so the principal was objectively nondiscriminatory, as defined above. Using terms in the economics literature, we can also say that the principal's hiring decisions showed no statistical discrimination or taste-based discrimination (32–35).

Nevertheless, the principal was judged as significantly more discriminatory against female applicants in the accept condition than in the reject condition, $t(446) = 6.00$, $P < 0.001$, $d = 0.57$. Specifically, in the accept condition, the principal was judged as discriminatory against females, $t(227) = -5.19$, $P < 0.001$, $d = 0.35$, compared with the midpoint of the scale; in the reject condition, the same principal was judged as discriminatory against males, $t(219) = 3.41$, $P = 0.001$, $d = 0.23$, compared with the midpoint of the scale.

To test our theory regarding the equality expectation, we recruited another nationally representative sample from the United States ($n = 154$), assigned them to the same conditions, and showed them the same hiring case. Instead of providing the composition of the accepted or rejected pool, we asked participants to indicate the ideal composition. In both the accept and reject conditions, the expected (ideal) proportion of women was close to equality ($M_s = 49.22\%$ and 47.01% ; medians = 50% and 50%). Equality was an unrealistic expectation given that women comprised only 20% of the candidate pool, but the expectation explains why the principal was judged as discriminatory in both the accept and reject conditions in the main study.

Study 1B investigated discrimination against underrepresented minorities (URMs). Participants ($n = 302$, recruited on Prolific, where most workers are from Organisation for Economic Co-operation and Development [OECD] countries) read a case about a human resources (HR) director who was making hiring decisions. We replicated the framing effect: judged discrimination against URM applicants was stronger in the accept condition than in the reject condition, $t(300) = 4.08$, $P < 0.001$, $d = 0.47$.

As in Study 1A, we measured the expected (ideal) composition, but we did so within the main study instead of using another sample. We found that the expected proportion of URMs was close to 50% in both the accept and reject conditions ($M_s = 43.4\%$ vs. 42.5% ; medians = 50% vs. 50%). We also found correlations between the expected composition and

Table 1. Study Overview

Study (context)	Main objective/ finding	Condition besides framing	Parameters					Judged discrimination		
			Min members	Maj member	Min% _C	Min% _A	Min% _R	JD _A mean (SD)	JD _R mean (SD)	
Study 1A (hiring)	Demonstrates the proposed framing effect		Women	Men	20%	20%	20%	5.64 (1.85)	4.53 (2.06)	
Study 1B (hiring)	Replicates the proposed framing effect		URMs	Non-URMs	10%	10%	10%	5.60 (1.71)	4.80 (1.68)	
Study 1C (dismissal)	Replicates the proposed framing effect		Men	Women	16.7%	16.7%	16.7%	5.49 (1.38)	5.01 (1.42)	
Study 1D (performer)	Replicates the proposed framing effect		Southerners	Northerners	20%	20%	20%	6.01 (1.97)	4.55 (1.96)	
Study 2 (layoff)	Demonstrates the robustness of the framing effect by showing that it exists regardless of objective discrimination	Objectively against Min Objectively against Maj	Southerners	Northerners	30%	10%	50%	8.07 (2.04)	6.57 (1.71)	
						50%	10%	5.51 (1.75)	4.36 (2.17)	
Study 3 (grader)	Demonstrates a behavioral consequence of the framing effect		Biden supporters	Trump supporters	25%	25%	25%	4.26 (1.28)	3.49 (1.40)	
Study 4 (refugees)	Tests and supports the flipside-neglect component of our theory by showing that judgment in the dual frame lies between the accept and the reject frames		Ayronians	Byronians	25%	25%	25%	JD _A 4.92 (1.43)	JD _{Dual} 4.30 (1.57)	JD _R 3.97 (1.61)
Study 5 (admissions)	Tests and supports the flipside-neglect component of our theory by showing the moderating effect of a flipside-thinking prompt	Control	From the east side	From the west side	25%	25%	25%	5.55 (1.37)	4.72 (1.38)	
		Flipside thinking						4.97 (1.29)	4.95 (1.30)	
Study 6 (committee)	Tests and supports the equality-expectation component of our theory by showing the moderating effect of manipulating equality	Unequal composition	Republicans	Democrats	30%	30%	30%	6.59 (1.74)	4.56 (1.99)	
		Equal composition						4.99 (0.67)	5.03 (0.81)	
Study 7 (committee)	Tests and supports the Equality-expectation component of our theory by showing the moderating effect of manipulating expectations	Expectedly equal	From the Red States	From the Blue States	36%	36%	36%	6.07 (1.69)	4.65 (1.78)	
		Expectedly unequal	From Rhode Island	From the other 49 US states				3.33 (2.18)	4.39 (2.04)	

Note: Min%_C, Min%_A, and Min%_R are the proportions of Min members among all candidates, the accepted candidates, and the rejected candidates, respectively. JD_A and JD_R are judged discrimination in the accept and reject conditions, respectively; higher values indicate stronger judged discrimination against Min members.

judged discrimination: In the accept condition, participants who expected a higher percentage of URMs perceived stronger discrimination against URMs, $r = 0.15$, $P = 0.062$; in the reject condition, participants who expected a higher percentage of URMs perceived stronger discrimination against non-URMs, $r = -0.16$, $P = 0.052$. Although the expectations were measured (not manipulated), and the correlations were only marginally significant, the findings suggest that the judgment of discrimination is related to expectations.

Study 1C replicated the framing effect and ruled out inattention and poor quantitative skills as alternative explanations. Like Study 1A, Study 1C concerned gender discrimination. To test the generalizability of our effect, Study 1C differed from Study 1A in that the Min members in Study 1C were men (rather than women), and the decision was about whether to keep or dismiss existing employees (rather than whether to hire or reject new applicants). Furthermore, we incentivized participants to make accurate judgments, asked them comprehension questions (before they judged discrimination), and tested their ability to infer the information on the flipside (after they judged discrimination).

Despite the differences, we again observed a significant framing effect: participants who saw how many men and women were kept (accept condition) perceived stronger discrimination against men than participants who saw how many men and women were dismissed (reject condition). The framing effect was significant whether we included all participants ($n = 306$, recruited from CloudResearch in the US), $t(303.79) = 2.98$, $P = 0.003$, $d = 0.34$, or only those who passed all comprehension questions and correctly inferred the information on the flipside (95.8% of all participants), $t(288.99) = 2.69$, $P = 0.008$, $d = 0.32$. The framing effect seems quite robust, is not limited to situations where the Min members are traditional subjects of discrimination (e.g., women), and cannot be explained by inattention or poor quantitative skills.

While the other studies used participants mainly from North America and Europe, Study 1D used participants from Asia ($n = 300$, recruited on Credamo in China) and replicated the framing effect, $t(298) = 6.44$, $P < 0.001$, $d = 0.74$, indicating that the effect is robust to cultural variations.

In Studies 1A to 1D, we designed the stimuli so that $\text{Min}\%_A = \text{Min}\%_R$ (i.e., the decision-maker exhibited no objective discrimination). According to our theory, the framing effect occurs even if $\text{Min}\%_A$ does not equal $\text{Min}\%_R$ (i.e., the decision-maker displays objective discrimination). Our theory also predicts that, on average, JD_A and JD_R will be greater if $\text{Min}\%_A < \text{Min}\%_R$ than if $\text{Min}\%_A > \text{Min}\%_R$. In other words, on average, participants should judge the decision-maker as more discriminatory against Min members if the decision-maker objectively discriminated against Min members than if the decision-maker objectively discriminated against Maj members.[‡]

Study 2 tested these predictions in a 2 (framing: accept vs. reject) \times 2 (objective discrimination: against Min vs. against Maj) \times 2 (subject design, using participants recruited on CloudResearch in the United States ($n = 606$)). As predicted, a 2 \times 2 ANOVA found a significant framing effect, $F(1, 602) = 71.26$, $P < 0.001$, $\eta^2 = 0.106$, a significant objective-discrimination effect, $F(1, 602) = 232.41$, $P < 0.001$, $\eta^2 = 0.279$, and no interaction effect, $F(1, 602) = 1.29$, $P = 0.257$. As the

[‡]These predictions can be derived formally as follows: According to our theory (Eqs. 1A and 1B), whether $\text{Min}\%_A$ is greater or smaller than $\text{Min}\%_R$ does not affect the difference between JD_A and JD_R , but it does affect the average of JD_A and JD_R : $(\text{JD}_A + \text{JD}_R)/2 = (\text{Min}\%_R - \text{Min}\%_A)/2 + k$, so the average of JD_A and JD_R will be greater if $\text{Min}\%_R > \text{Min}\%_A$ than if $\text{Min}\%_R < \text{Min}\%_A$. Since the framing effect pertains to the difference between JD_A and JD_R , and the overall judgment pertains to the average of JD_A and JD_R , the former does not depend on objective discrimination, while the latter does.

results shown in Fig. 1 reveal, judged discrimination is both distinct from and reflective of objective discrimination: It is distinct because it varies (as a result of framing) when objective discrimination is held constant. Also, it is reflective because it varies when objective discrimination varies.

Study 3 demonstrated a behavioral consequence of the framing effect with monetary incentives. Participants ($n = 304$, recruited on CloudResearch in the United States) were asked whether they supported Biden or Trump in the 2020 election. They learned that they would be writing an essay to explain their choice, but before writing the essay, participants learned that a research assistant (RA) would read their essay and would either pass or fail them, which would influence their earnings.

We told participants that the RA had already evaluated the essays of 48 workers, of whom 36 were Biden supporters (the majority) and 12 were Trump supporters (the minority); the RA had passed half the workers and failed half the workers. We then told participants in the [accept] {reject} condition that among the workers the RA had [passed] {failed}, 18 were Biden supporters and 6 were Trump supporters. The candidates in this study were the 48 workers the RA had evaluated, the Min members were the Trump supporters, and $\text{Min}\%_C = \text{Min}\%_A = \text{Min}\%_R = 25\%$. Participants were then asked (a) to evaluate whether the RA discriminated against either Trump or Biden supporters, and (b) to decide whether to keep this RA or replace the RA with a different RA.

Fig. 2A shows the judgment results. We replicated the framing effect: participants in the accept condition judged the RA as more discriminatory against Trump supporters (the Min group) than participants in the reject condition, $t(302) = 5.02$, $P < 0.001$, $d = 0.58$. The effect was significant among the Biden supporters, $t(152.43) = 5.39$, $P < 0.001$, $d = 0.81$, and marginally significant among the Trump supporters, $t(97) = 1.87$, $P = 0.065$, $d = 0.37$.

Fig. 2B shows the behavioral results. Whether participants wanted to change RAs depended on whether participants judged the current RA as discriminatory against those who had supported the participant's preferred presidential candidate or against those who had supported the other presidential candidate ($r = 0.477$, $P < 0.001$). Specifically, fewer Biden supporters in the accept condition than in the reject condition wanted to change RAs: 11.3% vs. 51.2%, $\chi^2(1, n = 181) = 34.15$, $P < 0.001$, $\phi = 0.43$. Conversely, more Trump supporters in

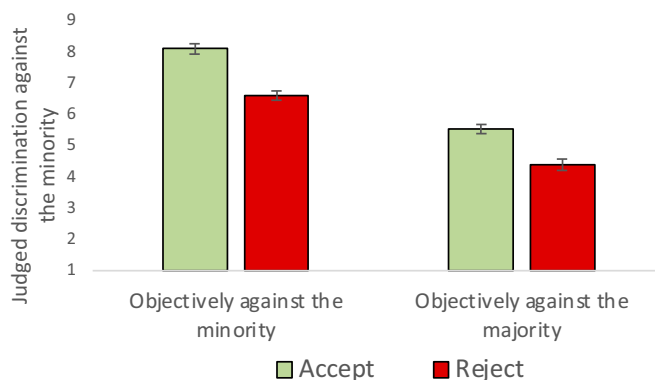


Fig. 1. Judged discrimination in Study 2. The decision-maker was judged as more discriminatory against the minority in the accept condition than in the reject condition, regardless of whether the decision-maker was objectively discriminatory against the minority or against the majority. Orthogonally, the decision-maker was judged as more discriminatory against the minority when the decision-maker was objectively discriminatory against the minority than when the decision-maker was objectively discriminatory against the majority. The error bars represent ± 1 SE.

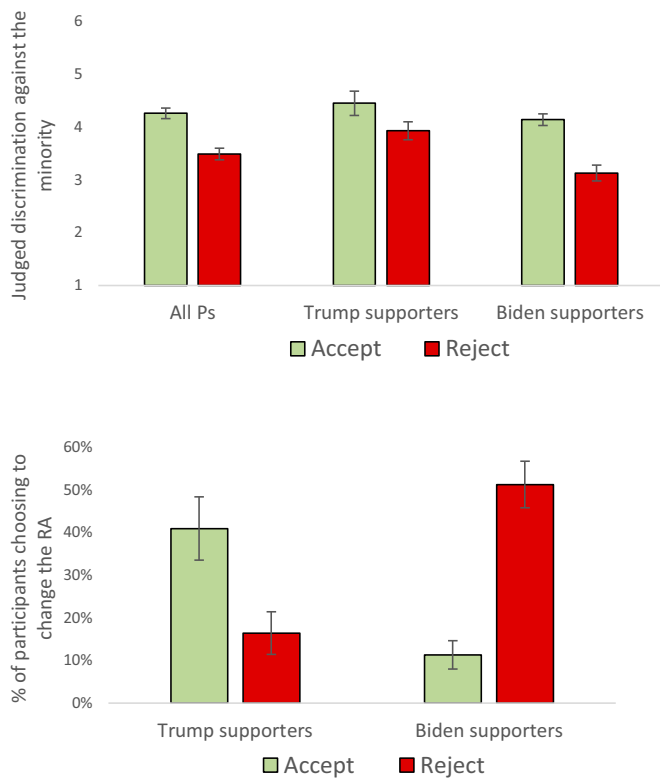


Fig. 2. (A) Judged discrimination in Study 3. The decision-maker (the RA) was judged as more discriminatory against the minority (Trump supporters) in the accept condition than in the reject condition, regardless of whether the participants themselves were Trump supporters or Biden supporters. The error bars represent ± 1 SE. (B) Choice behavior in Study 3. More Trump supporters in the accept condition than in the reject condition chose to change RAs, and more Biden supporters in the reject condition than in the accept condition chose to change RAs. The error bars represent ± 1 SE.

the accept condition than in the reject condition wanted to change RAs: 40.9% vs. 16.4%, $\chi^2(1, n = 99) = 7.43, P = 0.006, \phi = 0.27$ (Fig. 2B).

The results of Study 3 demonstrate that framing influences not only judgment but also choice. People may either expel a decision-maker or keep her, depending on whether they see who the decision-maker accepted or who she rejected.

The studies reported so far established the robustness of the framing effect and demonstrated a behavioral consequence. The remaining studies tested the two psychological underpinnings of our theory: flipside neglect in Studies 4 and 5, and the equality expectation in Studies 6 and 7.

Study 4 included three framing conditions: an accept condition, a reject condition, and a dual-frame condition, which included information about the compositions of both the accept and reject groups. If the framing effect found in the other studies is indeed due to neglect of the flipside information, then the judgment of participants in the dual-frame condition should lie between the judgment of participants in the other two conditions—and this is what we found.

Participants ($n = 448$, recruited on Prolific) read a case in which a rich country was deciding which refugees from a neighboring country to keep and which to deport. The refugee pool contained members of two fictional ethnic groups, Ayrionians (minority) and Byronians (majority); participants judged whether the rich country discriminated against either group. We observed a significant framing effect, $F(2, 445) = 14.58, P < 0.001, \eta^2 = 0.061$. Importantly, we found that judged discrimination in the dual-frame condition fell between judged

discrimination in the accept and reject conditions: relative to participants in the dual-frame condition, participants in the accept condition judged the rich country as significantly more discriminatory against Ayrionians (Min members), $P = 0.001$, and participants in the reject condition judged the same country as (marginally-significantly) more discriminatory against Byronians (Maj members), $P = 0.067$.

Study 5 more directly tested the role of flipside neglect by explicitly prompting some participants to calculate the flipside information before judging discrimination. If the framing effect is due to a failure to spontaneously consider information on the flipside, then the prompt should mitigate the framing effect. To test this prediction, Study 5 adopted a 2 (framing: accept vs. reject) \times 2 (control vs. flipside-thinking) between-subjects design. All participants ($n = 604$; recruited on Prolific) read a case in which 400 prospective students (100 from the city's east side, 300 from the city's west side) applied to a city college. The college admitted half the applicants and rejected the other half; participants in the [accept] {reject} condition learned that among the [admitted] {rejected} applicants, 50 were from the east side, and 150 were from the west side.

Then, participants received one of two prompts: a mere-thinking prompt (control) or a flipside-thinking prompt. The mere-thinking prompt said, "Take a moment to think about the information in the case." The flipside-thinking prompt was more specific: "Take a moment to think about the following: Given that 100 people from the east side and 300 people from the west side applied, and the college [admitted] {rejected} 50 from the east side and 150 from the west side, then how many from the east side and how many from the west side did the college [reject] {admit}?" Notably, the flipside-thinking prompt conveyed no new information; it simply asked participants to calculate the flipside information by themselves.

As predicted, a 2 \times 2 ANOVA on judged discrimination yielded a significant framing effect, $F(1, 600) = 15.13, P < 0.001, \eta^2 = 0.025$, no significant prompt effect, $F(1, 600) = 2.50, P = 0.114$, and a significant interaction, $F(1, 600) = 13.78, P < 0.001, \eta^2 = 0.022$. In the control (mere-thinking) condition, we replicated the framing effect, $F(1, 600) = 28.90, P < 0.001, \eta^2 = 0.046$, which corroborates prior findings that mere deliberation is insufficient to overcome a bias (36, 37). In the flipside-thinking condition, the framing effect disappeared, $F(1, 600) = 0.02, P = 0.900$. It seems that the framing effect occurs not because people fail to think, but because they fail to think correctly—they do not consider the flipside information unless explicitly prompted to do so. (The fact that flipside-thinking prompt did not flip the framing effect suggests that participants had already formed a judgment of the decision-maker in their given frame before receiving the prompt, and the prompt merely canceled out the effect of the original frame but did not override it.)

Study 6 and Study 7 tested the other component of our theory, the equality expectation. According to our theory, the framing effect occurs because people expect the composition of the constituent groups in their condition to be roughly equal (even though the composition of the constituent groups in the candidate pool is unequal). This theory predicts two moderators. First, if the composition of the constituent groups in the candidate pool is equal, the framing effect will disappear. (This prediction can be easily derived from Eq. 4, which implies that $JD_A = JD_R$ if $Min\%_C = 50\%$.) Second, the framing effect will weaken or even reverse if people expect the composition of the constituent groups to be unequal rather than equal. Study 6 tested the first prediction, and Study 7 tested the second prediction.

Study 6 adopted a 2 (framing: accept vs. reject) \times 2 (composition of the constituent groups in the candidate pool: unequal vs. equal) between-subjects design with participants recruited on CloudResearch in the United States ($n = 618$). As predicted, a 2 \times 2 ANOVA yielded a framing effect, $F(1, 614) = 75.60, P < 0.001, \eta^2 = 0.110$, a composition effect, $F(1, 614) = 24.76, P < 0.001, \eta^2 = 0.039$, and, most importantly, an interaction effect, $F(1, 614) = 81.61, P < 0.001, \eta^2 = 0.117$. Specifically, we replicated the framing effect when the constituent groups were unequally represented in the candidate pool, $F(1, 614) = 157.16, P < 0.001, \eta^2 = 0.204$, but not when the groups were equally represented, $F(1, 614) = 0.06, P = 0.811$.

Study 7 adopted a 2 (framing: accept vs. reject) \times 2 (constituent groups: expectedly equal vs. expectedly unequal) between-subjects design. All participants ($n = 608$, recruited on CloudResearch in the United States) read a case in which a US political leader was forming a bipartisan committee to discuss policy issues. One hundred Americans applied to serve on the committee; 36 were from a Min group, and 64 were from a Maj group. In the expectedly-equal conditions, the Min group was people from the Red States, and the Maj group was people from the Blue States. In the expectedly-unequal conditions, the Min group was people from Rhode Island, and the Maj group was people from the other 49 states. We assumed that participants would expect roughly equal representation when the two groups were from the Red States and the Blue states but would expect unequal representation (fewer people from the Min group) when the Min group was people from Rhode Island and the Maj group was people from the other 49 states. We verified this assumption in a separate pretest ($n = 202$).

All participants learned that the leader accepted 50 applicants and rejected 50 applicants. Participants in the [accept] {reject} condition learned that among the [accepted] {rejected} applicants, 18 were Min members and 32 were Maj members.

The results, displayed in Fig. 3, demonstrate the moderating role of expectations. A 2 \times 2 ANOVA on judged discrimination found no framing effect, $F(1, 604) = 1.32, P = 0.251$, a significant effect of the expectation manipulation, $F(1, 604) = 91.58, P < 0.001, \eta^2 = 0.132$, and a significant interaction, $F(1, 604) = 62.43, P < 0.001, \eta^2 = 0.094$. In the expectedly-equal conditions, we replicated the signature framing effect: the leader was judged as *more* discriminatory against Min members (people from the Red States) in the accept condition than in the reject condition, $F(1, 604) = 41.09, P < 0.001, \eta^2 = 0.064$. In the expectedly-unequal condition, the framing effect

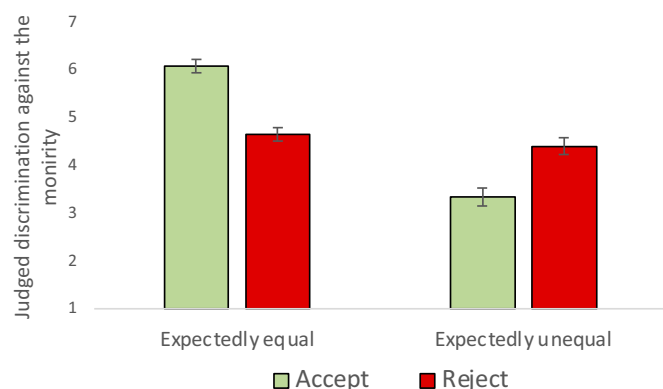


Fig. 3. Judged discrimination in Study 7. In the expected-equal condition (in which the minority candidates were from the Red States), the decision-maker was judged as more discriminatory against the minority in the accept condition than in the reject condition (i.e., the typical framing effect). In the expectedly-unequal condition (in which the minority candidates were from Rhode Island), the framing effect reversed. The error bars represent ± 1 SE.

reversed: the leader was judged as *less* discriminatory against Min members (people from Rhode Island) in the accept condition than in the reject condition, $F(1, 605) = 22.72, P < 0.001, \eta^2 = 0.036$. (*SI Appendix, SI Part 3* shows why the framing effect reversed rather than merely disappeared.) These findings highlight the role of the equality expectation in producing the framing effect.

Discussion

Discrimination is not just an objective fact but also a subjective judgment. We argue and show that the judgment of discrimination is malleable to framing manipulations, holding objective discrimination constant. We replicated this framing effect with participants from different parts of the world, with different types of discrimination, and with measures of judgment as well as behavior. The existence of the framing effect is evidence that judged discrimination is distinct from objective discrimination and is fallible.

We proposed a behavioral theory of how people judge discrimination: they neglect information on the flipside and they expect equality in the composition of the group they see (be it the accepted or rejected candidates). We tested and supported both components of the theory. In all our studies, we assumed that the constituent groups in the candidate pool were equally qualified, and we proposed (and found evidence that) observers' expectations are based on the equality expectation. In *SI Appendix, SI Part 3*, we delineate a more general theory that relaxes those assumptions. According to the general theory, the framing effect will arise even if one group is less qualified, and even if people in the accept and reject conditions hold different expectations, as long as the expected compositions are on average closer to equality than the compositions of the candidate pool.

Framing had a robust effect in our experiments, but we surmise that framing has an even greater effect in real life. In our experiments, participants in each frame had sufficient information to infer the information in the other frame, so they could resist the influence of framing (though they often did not). In real life, the public rarely has enough information to render unbiased assessments. For example, the public may easily observe how many men and women a firm currently employs, but it is hard to know how many men and women applied, so it is hard to infer the composition of the rejected group.

Our research suggests that the media and the decision-makers themselves can help the public judge discrimination more accurately by making relevant information accessible and nudging the public to use the information correctly. For example, when reporting on the hiring process of a firm, the media may want to report not only who the firm has hired, but also try to find out and report who the firm could have hired but did not and nudge the public to consider both sets of information when judging the discriminatory tendency of the firm. To facilitate accurate judgments by the public, the firm itself may also want to disclose (e.g., on its website) not just the demographic information of its current employees, but also the demographic information of all applicants or of the rejected applicants. These interventions may help the public identify real discrimination while minimizing false accusations.[§]

[§]To illustrate, consider two firms, A and B; each firm hired half of its job candidates and rejected the rest. In firm A, 30% of the hired candidates are women and 10% of the rejected candidates are women; in firm B, 40% of the hired candidates are women and 60% of the rejected candidates are women. If the public sees only the statistics of the hired candidates, they will likely judge firm A as more discriminatory against women. But if they also see the information on the reject side and know how to use the information, they will likely judge firm B as more discriminatory against women. This example illustrates that the information on the reject side and the knowledge to use it correctly could help the public not only realize which firm is less discriminatory, but also identify which firm is more discriminatory.

Data, Materials, and Software Availability. Anonymized [xlsx.] data have been deposited in [OSF] (https://osf.io/k8xsb/?view_only=2ef058bb8d84f0b9141c5145161488f) (38).

ACKNOWLEDGMENTS. The authors wish to thank the following individuals (in alphabetical order by last name) for their helpful suggestions on earlier versions

1. M. Bertrand, S. Mullainathan, Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* **94**, 991–1013 (2004).
2. A. W. Brooks, L. Huang, S. W. Kearney, F. E. Murray, Investors prefer entrepreneurial ventures pitched by attractive men. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4427–4431 (2014).
3. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16474–16479 (2012).
4. L. Quillian, D. Pager, O. Hexel, A. H. Midtbøen, Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10870–10875 (2017).
5. X. Li, C. K. Hsee, Free-riding and cost-bearing in discrimination. *Organ. Behav. Hum. Decis. Process.* **163**, 80–90 (2021).
6. M. T. Schmitt, N. R. Branscombe, T. Postmes, A. Garcia, The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychol. Bull.* **140**, 921–948 (2014).
7. E. A. Pascoe, L. Smart Richman, Perceived discrimination and health: A meta-analytic review. *Psychol. Bull.* **135**, 531–554 (2009).
8. National Science Foundation National Center for Science and Engineering Statistics, Women, minorities, and persons with disabilities in science and engineering: 2019. (Special report NSF 19-304) (2019). <https://ncses.nsf.gov/pubs/nsf19304/data>. Accessed 6 April 2022.
9. V. Narayanan, I'm an ex-Google woman tech leader and i'm sick of our approach to diversity! (2017). <https://medium.com/the-mission/im-an-ex-google-woman-tech-leader-and-i-m-sick-of-our-approach-to-diversity-17008c5fe999>. Accessed 6 April 2022.
10. Google, Representation at Google (2021). <https://diversity.google/annual-report/representation/>. Accessed 6 April 2022.
11. D. Kahneman, *Thinking, Fast and Slow* (Macmillan, 2011).
12. L. Brenner, D. Koehler, A. Tversky, On the evaluation of one-sided evidence. *J. Behav. Decis. Making* **9**, 59–70 (1996).
13. D. Hamilton, R. Gifford, Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *J. Exp. Soc. Psychol.* **12**, 392–407 (1976).
14. K. Fiedler, The tricky nature of skewed frequency tables: An information loss account of distinctiveness-based illusory correlations. *J. Pers. Soc. Psychol.* **60**, 24–36 (1991).
15. M. Bar-Hillel, The base-rate fallacy in probability judgments. *Acta Psychol. (Amst.)* **44**, 211–233 (1980).
16. D. Kahneman, A. Tversky, On the psychology of prediction. *Psychol. Rev.* **80**, 237–251 (1973).
17. M. Welsh, D. Navarro, Seeing is believing: Priors, trust, and base rate neglect. *Organ. Behav. Hum. Decis. Process.* **119**, 1–14 (2012).
18. B. Enke, What you see is all there is. *Q. J. Econ.* **135**, 1363–1398 (2020).
19. J. Koehler, M. Mercer, Selection neglect in mutual fund advertisements. *Manage. Sci.* **55**, 1107–1121 (2009).
20. K. Fiedler, Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychol. Rev.* **107**, 659–676 (2000).
21. F. Leclerc, C. Hsee, J. Nunes, Narrow focusing: Why the relative position of a good in its category matters more than it should. *Mark. Sci.* **24**, 194–205 (2005).
22. M. Rabin, G. Weizsäcker, Narrow bracketing and dominated choices. *Am. Econ. Rev.* **99**, 1508–1543 (2009).
23. D. Read, G. Loewenstein, M. Rabin, Choice bracketing. *J. Risk Uncertain.* **19**, 171–197 (1999).
24. A. Tversky, D. Kahneman, The framing of decisions and the psychology of choice. *Science* **211**, 453–458 (1981).
25. E. Shafir, Choosing versus rejecting: Why some options are both better and worse than others. *Mem. Cognit.* **21**, 546–556 (1993).
26. E. Shafir, I. Simonson, A. Tversky, Reason-based choice. *Cognition* **49**, 11–36 (1993).
27. E. Fehr, K. Schmidt, A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
28. R. Subrahmanian, Gender equality in education: Definitions and measurements. *Int. J. Educ. Dev.* **25**, 395–407 (2005).
29. M. Bazerman, S. White, G. Loewenstein, Perceptions of fairness in interpersonal and individual choice situations. *Curr. Dir. Psychol. Sci.* **4**, 39–43 (1995).
30. K. Hugenberg, G. V. Bodenhausen, M. McLain, Framing discrimination: Effects of inclusion versus exclusion mind-sets on stereotypic judgments. *J. Pers. Soc. Psychol.* **91**, 1020–1031 (2006).
31. L. T. Phillips, S. Jun, Why benefiting from discrimination is less recognized as discrimination. *J. Pers. Soc. Psychol.* **122**, 825–852 (2022).
32. G. Becker, *The Economics of Discrimination* (University of Chicago Press, 1957).
33. E. S. Phelps, The statistical theory of racism and sexism. *Am. Econ. Rev.* **62**, 659–661 (1972).
34. J. Bohren, K. Haggag, A. Imas, D. Pope, Inaccurate Statistical Discrimination: An Identification Problem. (2019) <https://nber.org/papers/w25935>. Accessed 28 October 2022.
35. J. Bohren, P. Hull, A. Imas, Systemic Discrimination: Theory and Measurement. (2022). <https://www.nber.org/papers/w29820>. Accessed 28 October 2022.
36. M. Lawson, R. Larrick, J. Soll, Comparing fast thinking and slow thinking: The relative benefits of interventions, individual differences, and inferential rules. *Judgm. Decis. Mak.* **15**, 660–684 (2020).
37. X. Li, C. Hsee, The psychology of marginal utility. *J. Consum. Res.* **48**, 169–188 (2021).
38. C. K. Hsee, X. Li, A framing effect in the judgment of discrimination. OSF Home. https://osf.io/k8xsb/?view_only=2ef058bb8d84f0b9141c5145161488f. Deposited 9 August 2022.

of this article: Alyssa Eldridge, Reid Hastie, Thomas Hsee, Alex Imas, Josh Klayman, Derek Koehler, Sendhil Mullainathan, Anuj Shah, Jack Soll, Shu Wang, Minwen Yang, Frank Yu, and participants in George Wu's lab at Chicago Booth. X.L. was a PhD student at Chicago Booth when the project commenced, and C. H. a visiting professor at CKGSB when the project concluded. Research support for this project was provided by Chicago Booth.