

(Not) Getting What You Deserve: How Misrecognized Evaluators Reproduce Misrecognition in Peer Evaluations*

Mabel Abraham
Columbia University
mabel.abraham@columbia.edu

Tristan L. Botelho
Yale University
tristan.botelho@yale.edu

James T. Carter
Cornell University
james.carter@cornell.edu

American Sociological Review, 2025

Final version: <https://journals.sagepub.com/doi/10.1177/00031224251318051>

Abstract

In most evaluation systems—such as those governing the allocation of prestigious awards—the evaluator’s primary task is to reward the highest quality candidates. However, these systems are imperfect; top performers may not be acknowledged and thus be underrecognized while low performers may receive unwarranted recognition and thus be overrecognized. An important feature of many evaluation systems is that people alternate between being candidates and being evaluators. How does experiencing misrecognition as a candidate affect how people subsequently evaluate others? We develop novel theory that underrecognition and overrecognition lead people to reproduce those experiences when they are evaluators. Across three studies—a quasi-natural experiment and two pre-registered, multistage experiments, we find that underrecognized evaluators are less likely to grant recognition to others—even to the highest performing candidates. Conversely, overrecognized evaluators are more likely to grant rewards to others—even to the lowest performing candidates. Whereas underrecognized evaluator behavior is driven by perceptions that their experience was unfair, overrecognized evaluator behavior is driven by the distinct informational cues they glean on how to evaluate others. Thus, in evaluation processes where people oscillate between being the evaluated and being the evaluator, we show how and why seemingly innocuous initial inefficiencies are reproduced in subsequent evaluations.

Keywords: evaluations, status, inequality, stratification

* Authors are listed in alphabetical order and contributed equally. We gratefully acknowledge Joel Brockner, Vanessa Burbano, Emilio Castilla, Jen Dannals, Stephan Meier, Ray Reagans, Amanda Sharkey, Mario Small, and Ezra Zuckerman for helpful comments and conversations. We also thank participants at the Columbia Business School Management Division Seminar and Yale University Sociology Seminar for helpful comments. All data and code can be accessed at https://osf.io/h6nfr/?view_only=86cb1887abab43d2a518e321549f34aa.

Evaluation processes are central to societal and organizational decisions governing the allocation of resources and rewards to individuals (Blau 1964; Homans 1951; Kim and DellaPosta 2022; Lamont 2012; Sharkey and Kovács 2017). This is particularly evident in professional contexts, where evaluations determine access to work opportunities (Mobasserri 2019; Weisshaar 2017) and who is recognized with prestigious awards (Bowers and Prato 2018; Jensen, Twardawski, and Younes 2021; Merton 1968). The goal of most evaluative processes, or at least their ideal, is to identify and recognize only the highest-quality candidates based on merit¹—namely a candidate’s performance and competencies (Alon and Tienda 2007; Castilla 2008; Castilla and Benard 2010; McNamee and Miller 2009; Mijs 2018; Scully 2000).

However, evaluative outcomes often deviate from this meritocratic ideal for a host of reasons, including that it is often difficult to objectively discern candidate performance (Correll, Benard, and Paik 2007; Lynn, Podolny, and Tao 2009; Paik et al. 2023). Instead, factors unrelated to candidate quality, such as their status or demographic attributes, are factored into evaluations, unduly benefitting certain candidates in ostensibly merit-based evaluations (Bowers and Prato 2018; Castilla 2008; Dobbin, Schrage, and Kalev 2015). Additionally, the design and structure of evaluative processes—including the degree to which evaluative criteria are formalized—shapes whether these processes can effectively identify and recognize top-performers (Botelho and Abraham 2017; Correll et al. 2020; Lucas et al. 2021; Rivera and Tilcsik 2019).

¹ There are many motives for evaluation processes to strive to reserve recognition for top candidates—and not recognize lower performers. Examples of such motives range from employers preferring to hire the most talented people to venture capitalists looking to fund the most promising ventures (Greenberg and Mollick 2017; Huang and Pearce 2015). In the case of prestigious awards and accolades, ensuring that only top-quality candidates be recognized preserves the value of these sources of recognition by maintaining the link between quality and recognition (Lynn, Podolny, and Tao 2009). Thus, we refer to the common motive for recognition to be reserved for high-quality candidates to mean that it goes only to top performers and not to lower performers.

There are clear consequences for individuals who experience unmeritocratic evaluative outcomes (i.e., those who do not receive the expected recognition or rewards). This phenomenon is commonly referred to as *misrecognition*, which encompasses both *underrecognition* and *overrecognition* (e.g., Kim and King 2014). In the case of *underrecognition*, high-quality individuals do not receive the recognition or rewards typical for their level of performance or quality, leading to subsequent disadvantages for that person. A classic example is Merton's (1968) discussion of the "41st chair," illustrating how scientists who just miss receiving a prestigious award, and thus were nearly identical in quality to the awardees, do not go on to enjoy the same future career benefits (see also Azoulay, Stuart, and Wang 2013). Conversely, *overrecognition* occurs when lower-quality candidates unexpectedly gain recognition and rewards, resulting in relative advantages over their comparable peers. A consequential and timely example of this phenomenon is college legacy admissions, where applicants benefit from preferential treatment due to familial connections to an institution, an advantage that is decoupled from academic merit (Espenshade and Radford 2009; Hurwitz 2011). Overrecognition similarly emerges in organizational contexts (Bond 2020). These instances of misrecognition highlight challenges inherent in evaluative processes that directly impact the resources a person later receives.

In this paper, we shift from examining the implications of misrecognition for individuals experiencing these evaluative outcomes to focusing on the downstream implications of misrecognition for how these individuals *subsequently evaluate others*. Some evaluative processes involve a distinct set of evaluators, such as third-party evaluations by experts or ranking institutions (Bowers and Prato 2018; Campanario 1998; Ody-Brasier and Sharkey 2019; Sharkey and Kovács 2017; Siler, Lee, and Bero 2015), but individuals frequently oscillate

between being evaluated and acting as the evaluator—across contexts and often even in the same evaluation system. This dynamic is notably observed in peer review systems prominent in science (Campanario 1998; Siler et al. 2015), performance evaluations in organizational contexts (Brett and Atwater 2001), and in evaluations on digital platforms (Botelho 2024; Klapper, Piezunka, and Dahlander 2024). Identifying whether and how people’s prior experiences with misrecognition affect their subsequent evaluative behavior is crucial for understanding when and why these types of evaluation processes might fail in recognizing top performers.

Drawing on—and bridging—theories related to evaluations, fairness, equity, and role-fulfillment, we propose that misrecognition among evaluators directly affects how they later allocate recognition and rewards to others. Effectively fulfilling the expectations associated with being an evaluator necessitates that evaluators understand *how to assess candidates*. Generally people strive to satisfy the expectations associated with their roles (Abelson 1981; Biddle 1986; Cialdini and Goldstein 2004; Goode 1960), and this motive tends to be even stronger for evaluators who serve as gatekeepers (Fini et al. 2022; Lamont 2012). Considering the widely-held belief in fair and merit-based outcomes (Lerner 1980; Mijs 2018) and the fact that most evaluation processes similarly espouse merit-based aims and criteria (Alon and Tienda 2007; Castilla 2008; McNamee and Miller 2009; Scully 2000), it seems plausible that evaluators would simply prioritize objective information about candidate quality in their assessments whenever possible. In other words, evaluators would generally recognize top performers—and not lower performers—irrespective of their own prior experiences having been evaluated.

However, we theorize that evaluators who have experienced misrecognition—or even simply believe they have been misrecognized—may be more apt to subsequently produce unmeritocratic evaluative outcomes. Specifically, experiences of misrecognition will lead

evaluators to subsequently reproduce misrecognition in their evaluations of others—such that underrecognized evaluators are *less* apt to grant recognition and overrecognized evaluators are *more* apt to grant recognition to others than similarly performing but correctly recognized peers. Individuals who have experienced any type of misrecognition have encountered an evaluative outcome that starkly contradicts the belief that recognition is reserved for top performers, thus challenging the assumed meritocratic nature of evaluation processes. For instance, the experience of underrecognition—namely strong performers who did not receive due recognition—works in direct opposition to the principle that only the highest-quality candidates should be acknowledged. Evaluators who have experienced misrecognition may therefore be influenced by their own dissonant experience, such that they reproduce their experiences of misrecognition in their subsequent evaluations of others.

We further propose that there are two distinct mechanisms underlying the reproduction of misrecognition that are a function of the type of misrecognition experienced—both stemming from the notion that these experiences inherently violate the general principle of meritocracy. First, experiences of misrecognition are apt to be perceived as inequitable or unfair (Adams 1965; Cook and Hegtvædt 1983). This perception of unfairness should manifest even among those overrecognized but is apt to be particularly salient among those disadvantaged by the unfairness—namely the underrecognized evaluators. Thus, we propose that the *perceived unfairness* of misrecognition will be the primary driver of underrecognized evaluators' subsequent behavior, leading them to underrecognize others.

Second, misrecognition also provides distinct information about how evaluation processes work that may shape how these individuals later evaluate others. In fact, people often rely not only on general information about how to behave or fulfill their role in a given context

but also on proximate environmental cues and personal experiences (Abelson 1981; Golden-Biddle and Rao 1997). An evaluator's experience of misrecognition provides additional—likely salient—informational cues about the appropriate evaluative criteria, which may lead them to reproduce their own experiences in their later evaluations of others. While all experiences of misrecognition provide informational cues on how to evaluate others, we reason that these *informational cues* will be an especially potent driver of overrecognized evaluators' subsequent behavior, leading them to overrecognize others.

To test our theoretical arguments, we pair a quasi-natural experiment in the field with two pre-registered, multistage online survey experiments. In our field study (Study 1), we observe peer evaluations by investment professionals on a private online platform that brings these professionals together to exchange and evaluate investment opportunities. In this setting, the allocation of recognition can be considered as good as random, such that we observe top (and equally) performing investment professionals who were correctly recognized with an award and equally deserving others who were not given the award (underrecognized). We examine how experiencing underrecognition affects professionals' subsequent evaluative behavior by comparing the way these professionals evaluated peers on this platform *before* and then *after* not receiving the award.

In Study 2 and Study 3, we aim to replicate our findings from Study 1 and extend these in two primary ways: By examining not only the effect of underrecognition but also of overrecognition on how evaluators allocate recognition to others and by exploring potential mechanisms underlying the effects of misrecognition. For these two latter studies, we recruited participants from an online sample and hired them to complete short jobs. They first completed an aptitude test that established whether they were a high or low performer. We then manipulated

experiences of misrecognition by randomly assigning people to receive recognition (or not) and had them subsequently evaluate responses to a similar aptitude test from an ostensible peer.

Our field study provides initial evidence that a professional's prior experience of underrecognition subsequently changes their behavior as an evaluator. Before the recognition event, investment professionals—both those who are later recognized (i.e., correctly recognized) and those who are not later recognized (i.e., underrecognized)—submitted similar ratings when evaluating peers' recommendations. However, after the award, underrecognized professionals submitted lower ratings relative to those they had previously submitted and relative to those submitted by correctly recognized professionals. Study 2 replicates this effect of underrecognition, providing causal evidence that underrecognized evaluators were less likely to grant recognition to others—even those who they indicated were the highest performing candidates. Study 2 also extends our findings to causally identify the effect of overrecognition: We show that overrecognized evaluators were more than twice as likely as their correctly recognized counterparts to grant recognition to others—even to those they indicated as the poorest performers.

Study 2 and Study 3 also test our theorized mechanisms. Consistent with our predictions, our results reveal there are two distinct mechanisms underlying the effects of misrecognition. For underrecognized evaluators, we find that perceptions that they themselves were evaluated unfairly drives them to grant recognition to others at a lower rate. Though overrecognized evaluators perceived the way they were evaluated as similarly unfair, this is not a driver of their subsequent evaluative behavior. Rather, overrecognized evaluators' subsequent behavior is driven by the informational cues they glean about how they *ought* to evaluate others from their own experiences having been evaluated. We do not find that the same informational cues have a

substantive influence on how underrecognized evaluators subsequently grant recognition to others.

This paper thus advances our understanding of the effects of recognition and meritocracy in evaluation processes by bridging sociological and related research on evaluations and theories of equity, procedural fairness, and role fulfillment. Our results show how experiencing unmeritocratic outcomes (i.e., instances of misrecognition) are often reproduced in evaluation systems where people move between being candidates and being evaluators.

THE EFFECTS OF MISRECOGNITION ON SUBSEQUENT EVALUATIONS

Evaluators are at the heart of every evaluation process, serving as gatekeepers who control access to vital resources and opportunities (Bian et al. 2022; Fini et al. 2022; Lamont 2012; Rivera 2015; Smith 2005); therefore, the extent to which they adhere to meritocratic ideals dictates the efficacy of these processes. Indeed, decades of research confirms that people believe that the allocation of resources and rewards is—and ought to be—fair and reflective of meritocratic processes (Jost and van der Toorn 2012; Lerner 1980), and these convictions have only strengthened over time (Mijs 2018). It is thus plausible that these beliefs primarily dictate how evaluators engage in their assessments of others, such that they grant recognition only to those who are truly deserving based on their performance or merit. However, existing research documents notable variation in people’s adherence to these principles across professional contexts—even in cases where evaluators had concrete, unambiguous, and objective information about a candidate’s performance or quality on which to base their assessments (Botelho and Abraham 2017; Bowers and Prato 2018; Dobbin et al. 2015; Foschi 2000; Mobasserri 2019; Pedulla 2016).

Consistency in applying meritocratic principles across evaluators is requisite for any evaluative process to successfully identify—and reserve rewards for—top candidates. Understanding the set of factors that lead to variations in evaluative behavior is therefore critical, and existing research points to two distinct factors associated with such variation. The first factor is the design and structure of evaluation processes (Botelho and Abraham 2017; Correll et al. 2020; Rivera and Tilcsik 2019). For example, evaluative behavior varies more significantly across evaluators, often leading to a higher propensity for unmeritocratic outcomes, when there are constraints on the evaluators’ ability to focus on each candidate (e.g., because of time/cognitive constraints) (Botelho and Abraham 2017; Bowers and Prato 2018; Lynn et al. 2009; Simcoe and Waguespack 2010). Macro properties of evaluation processes may also have widespread effects on observed inequality, such as whether evaluative outcomes result in fine-grained distinctions between candidates (e.g., Accominotti, Lynn, and Sauder 2022). Similarly, less formalized processes, which offer greater discretion to evaluators, also promote inconsistencies in evaluative outcomes across evaluators (Anderson and Tomaskovic-Devey 1995; Dobbin 2009; Elvira and Graham 2002).

The second factor relates to characteristics of the evaluator. Research in this tradition has mostly focused on how certain demographic characteristics of evaluators—such as race or gender—are associated with variations in how they make merit-based resource allocation decisions. For example, members of marginalized groups, including women and racial minorities, often demonstrate a higher propensity to base evaluations on objective quality rather than on candidate characteristics unrelated to quality (e.g., gender, race) and thus tend to be more meritocratic (Abraham 2017; Cohen, Broschak, and Haveman 1998; Cohen and Huffman 2007; Yang and Aldrich 2014).

Because individuals frequently serve as both candidates and evaluators—across contexts but often even in the same evaluation system—we propose a third factor that is apt to drive variation in evaluative behavior: People’s prior experiences with having been evaluated. Recent sociological research generally supports this notion, indicating that both prior positive (Botelho and Gertsberg 2021) and prior negative (Castilla and Ranganathan 2020) evaluations can lead individuals to adopt fairer and less biased approaches in their subsequent evaluations. Building on this foundation, we propose that it is not simply the valence of how an evaluator was previously evaluated that shapes their subsequent evaluative behavior. Rather, experiences of misrecognition—whereby evaluators have previously received either less or more recognition than expected for their level of performance—will also significantly affect how people subsequently evaluate others. Instances of misrecognition directly contradict the central tenet of meritocratic norms that underlie most evaluative processes: That evaluation processes are centered on identifying and recognizing only top performing candidates as determined by their accomplishments or contributions (Castilla 2008; Castilla and Benard 2010; McNamee and Miller 2009; Scully 2000). Therefore, misrecognized evaluators may be more inclined to deviate from meritocratic principles in their own evaluative roles, either consciously or subconsciously, such that they misrecognize others.

Importantly, how misrecognition affects subsequent evaluative behavior will depend on the specific type of misrecognition—namely whether an evaluator has experienced underrecognition or overrecognition. Unlike those who have received the recognition expected, we argue that evaluators who have experienced misrecognition—or even those who simply perceive that they were misrecognized—will *reproduce their experiences of misrecognition* in how they subsequently grant recognition to others. Underrecognized evaluators are top

performers who did not receive the recognition or rewards typical for their level of performance; thus they will reproduce underrecognition by granting recognition to others at a lower rate. By contrast, overrecognized evaluators are lower performers who were nonetheless recognized and will thus grant recognition more freely, such that they overrecognize others.

Hypothesis 1a: Underrecognized evaluators will be less likely to grant recognition to others—even those they deem to be top performers—than will evaluators who performed similarly but were correctly recognized.

Hypothesis 2a: Overrecognized evaluators will be more likely to grant recognition to others—even those they deem to be poor performers—than will evaluators who performed similarly but were correctly recognized.

DRIVERS OF THE REPRODUCTION OF MISRECOGNITION: (UN)FAIRNESS AND INFORMATIONAL CUES

We draw on broader theories related to equity, procedural fairness, and role-fulfillment to highlight the two main theoretical reasons why evaluators who have experienced misrecognition will reproduce such patterns in their subsequent assessments of others. Both reasons stem from the notion that both types of experiences with misrecognition violate the general expectation that recognition is granted only to top performers: These experiences are apt to both invoke perceptions of unfairness and provide distinct informational cues not available to those who have been correctly recognized. Misrecognition may thus affect how evaluators later grant recognition to others either because of the perceived (or inherent) unfairness they experienced and/or by providing them with informational cues about how to evaluate others. In the following sections,

we explore both underlying mechanisms to consider when each is most apt to drive the reproduction of misrecognition.

Perceptions of (Un)Fairness among Misrecognized Evaluators and Their Subsequent Evaluations of Others

Those who did not receive the recognition or reward consistent with their performance are apt to perceive these experiences as unfair, which may drive how these misrecognized evaluators subsequently assesses others. Research in equity theory first brought attention to situations where rewards and contributions do not align (Adams 1965). Specifically, in cases where social comparison is possible—such as in workplaces—people tend to compare whether their outcomes (e.g., pay, promotion) are similar to those of coworkers who are analogous to them in terms of their contributions or merit (e.g., skills, performance) to determine whether these outcomes are equitable (Carrell and Dittrich 1978; Cook and Hegtvedt 1983). For example, underrecognized workers who did not receive a top performance rating have been shown to be more likely to exit their firm than similar peers who received a top rating (Bond 2024).

Experiences of misrecognition are likely to elicit perceptions of inequity, or unfairness, even in the absence of opportunities for direct social comparison. In fact, such violations of expectations are a key predictor of whether people interpret outcomes and processes as unjust (Cropanzano and Ambrose 2001), with people constantly making judgements about whether outcomes are fair or not (Lind 2001). For example, perceived unfairness is the most common cause of interpersonal sabotage and deviance in the workplace (Ambrose, Seabright, and Schminke 2002). Since experiences of misrecognition do not align with the meritocratic expectations that rewards ought to be granted only to the top-performing or best candidates, they are likely to be viewed as unfair. Specifically, misrecognized individuals are apt to perceive their

outcome as procedurally unfair; that is, they are likely to see the process used to make decisions about the allocation of recognition as unfair (Konovsky 2000; Lind and Tyler 1988).

A core focus of the current research is understanding how and under what conditions experiences of misrecognition—and by extension unfair outcomes—influence the way evaluators subsequently allocate recognition and rewards to others. Experiencing unfairness often shapes individuals' behavior, typically eliciting negative (or undesirable) actions (Ambrose et al. 2002) aimed at rectifying their experiences of inequity (Adams 1965; Gaucher et al. 2010; Lerner 1980). For example, a large body of research has tested equity theory in both the lab and the field, examining how job satisfaction and subsequent work behaviors are influenced by having respondents recall how their compensation compares to their peers (e.g., Sweeney 1990) or by experimentally manipulating inequity in resources (Carrell and Dittrich 1978; Shaw and Olson 2012). This research consistently reveals that people feel dissatisfaction or discomfort with inequity and try to reduce it by adjusting their contributions to align with the rewards and recognition they have received. In a study of professional baseball and basketball players, for instances, Harder (1992) found that underrecognized athletes (i.e., those who were under rewarded for their performances) were less cooperative and more selfish. This point is supported by research that has demonstrated the ways in which perceived unfairness triggers loss aversion concerns, such that people are intrinsically motivated to avoid such situations (Lieberman, Idson, and Higgins 2005; Zhou and Wu 2011).

Although unfairness is often associated with experiences that are disadvantageous (i.e., instances of negative inequity such as underrecognition), advantageous outcomes where the reward exceeds what is warranted by merit (i.e., such as overrecognition) also constitute a form of unfairness, termed “positive inequity” (Adams 1965; Brockner 1985; Cook and Hegtvedt

1983; Rivera and Tedeschi 1976; Scheer, Kumar, and Steenkamp 2003). One key distinction is that people generally welcome favorable outcomes, irrespective of how they get them (van den Bos et al. 1998). Also, unlike negative inequity, positive inequity typically does not evoke distress or displeasure (Rivera and Tedeschi 1976; Weiss, Suckow, and Cropanzano 1999), nor does it lead to a sense of perceived loss like negative inequity (Kidd and Utne 1978; Walster and Piliavin 1972). In fact, many of the undesirable behaviors linked to unfairness are based on studies focused on instances of negative inequity (e.g., Cohen-Charash and Spector 2001; Gaucher et al. 2010; Harder 1992; Yang et al. 2014).

By contrast, positive inequity tends to elicit feelings of guilt (Baumeister, Stillwell, and Heatherton 1994). Consistent with this finding, and contrary to underpaid employees who tend to reduce their productivity, overpaid workers often increase their output, although the evidence for this positive effect is mixed (Adams and Rosenbaum 1962; Lawler 1968; see Carrell and Dittrich 1978 for an example of a null effect). In other words, positive inequity seems to lead to a greater investment in the inequality producing system, whereas negative inequity tends to have the opposite effect.

Although evaluators who have experienced either underrecognition or overrecognition are apt to perceive their experiences as less fair than those who have been correctly recognized, the propensity for perceived unfairness to drive evaluators' subsequent behavior is likely asymmetrical; therefore, perceived unfairness is most apt to underlie underrecognized evaluator behavior since these individuals experienced an instance of negative inequity. Research has consistently shown that the adverse impact of negative inequity is stronger than any positive effects of positive inequity (Austin and Walster 1974; Brockner et al. 1994). This research thus suggests that perceived unfairness will be an especially strong driver of evaluator behavior

among those who have previously experienced underrecognition, but not for those who have been overrecognized.

Hypothesis 1b: The lower likelihood for underrecognized evaluators to grant recognition will be driven by perceptions of unfairness.

Informational Cues Available to Misrecognized Evaluators and Subsequent Evaluations

Though perceived (un)fairness may not directly drive overrecognized evaluator behavior, the positive inequity associated with overrecognition is apt to make these evaluators especially attentive to the prevailing evaluative criteria. Evaluators are responsible for making a judgement about a candidate or pool of candidates, often to determine who ought to be selected and rewarded (Rivera 2015; Smith 2005). Understanding how to evaluate candidates—most importantly, the evaluative criteria defining what is valued—is thus requisite for effectively fulfilling one’s role as an evaluator. More generally, people tend to be motivated to fulfill the responsibilities and expectations associated with the roles and positions they occupy (Abelson 1981; Biddle 1986; Goode 1960), and this is especially true for evaluators who tend to value the importance of their gatekeeper position (Bourdieu 1993; Lamont 2012). Given that positive inequity leads people to make even greater investments in the systems that have advantaged them, it stands to reason that overrecognized evaluators will be particularly focused on available cues about the appropriate way to evaluate others.

Evaluators may draw inferences about how to assess candidates from multiple distinct sources of information. The first of these is rooted in overarching meritocratic beliefs suggesting that evaluations are intended to identify and reward only the best performers. Though offering a reasonable and general guidepost, knowledge of meritocratic norms alone does not typically provide sufficient insight for how an evaluator should assess candidates (Alon and Tienda 2007). For instance, an evaluator needs to know what constitutes a candidate being the “best” or a “top

performer” in a specific context. A more relevant and precise source of information therefore comes from the evaluative criteria defined by the specific evaluation process. In more structured or formalized processes, these criteria are clearly defined and offer unambiguous rules and procedures governing how evaluators are expected to judge candidates (Elvira and Graham 2002; Reskin 2000; Sutton et al. 1994).

In many cases, however, these criteria include ambiguous elements, or at least, they allow evaluators the latitude in determining how to assess others (Lamont 2012; Rahman 2021; Veen, Barratt, and Goods 2020). Moreover, evaluative criteria are sometimes intentionally designed to be ambiguous. In many professional contexts, for example, less formalized criteria are used to afford managers greater discretion when doling out benefits (Gallus and Frey 2016). Evaluators are thus often left to determine on their own how to effectively fulfill their role. Hence, when criteria are left ambiguous or less formalized, there is greater variance in how evaluators assess candidates (Castilla 2008; Uhlmann and Cohen 2005).

Prior experiences being evaluated can serve as an additional valuable guide for evaluators to determine how to effectively judge candidates. People generally rely on their own relevant experiences for understanding how they should behave in a given context (Biddle 1986; Golden-Biddle and Rao 1997). This implies that evaluators may rely on both the overarching evaluative criteria—which typically aligns with meritocratic principles that the highest-quality candidates should be recognized—and their prior experiences of being evaluated to inform how to effectively evaluate others. For example, recent qualitative research on how managers understand and define merit suggests that people draw upon their broad set of prior experiences having been evaluated throughout their careers and across organizations in order to determine how to effectively evaluate their current employees (Castilla and Ranganathan 2020). The most

significant insights are likely drawn from experiences *within the same system*, given that the specific context and norms of this environment are most informative. Consistent with this line of reasoning, the way managers and executives enact their roles is heavily influenced by their experiences and observations within the specific organizational setting (Golden-Biddle and Rao 1997).

Evaluators' prior experiences with being evaluated not only offer them information in cases where evaluation criteria are ambiguous, but also provide misrecognized evaluators novel insights about how to judge candidates even when criteria are clear. Experiences of misrecognition are apt to violate the stated evaluative criteria since these tend to prioritize recognition being reserved for the best candidates. When a person's prior evaluative outcomes were correct, such that they received the warranted recognition given their performance, their experience having been evaluated reaffirms the expectation that recognition should be awarded only to the best or strongest performers. When top performers are duly recognized, for instance, it understandably validates their a priori meritocratic assumptions. Conversely, experiences of misrecognition provide information that directly violates expectations that merit governs evaluative outcomes. In the case of underrecognition—such as Merton's "41st chair"—a strong performer who does not receive recognition despite their high performance may begin to question the validity of their prior assumptions about who ought to be recognized. Similarly, individuals who are overrecognized, who receive recognition despite their lower performance, might also develop a skewed perception of the appropriate way to assess others.

While all misrecognized evaluators gain access to distinct informational cues stemming from their experience having been evaluated, we reason that overrecognized evaluators will be especially attentive and reactive to these informational cues. As poorer performers who were still

recognized, overrecognized evaluators are apt to infer that recognition is *not* exclusively reserved for top performers. Or more generally, they may infer that there are other factors at play in the determination of who receives recognition. Because overrecognized evaluators have themselves been advantaged, they may be particularly apt to seek out or be attentive to additional information in their environment to provide guidance on how they should recognize others. This point is consistent with classic research in equity theory that articulates a person will increase their inputs (e.g., attention, effort) to account for their overrecognition (e.g., Adams and Rosenbaum 1962). Thus, attending to these informational cues on how to evaluate others would lead overrecognized evaluators to subsequently grant recognition to others more freely, such that they are more likely to overrecognize those whom they evaluate.

Hypothesis 2b: The higher likelihood for overrecognized evaluators to grant recognition will be driven by the informational cues they have available.

DATA AND METHODS

Despite the prevalence of experiences of misrecognition, collecting appropriate observational data to test our theory about the downstream effects of misrecognition on subsequent evaluative behavior is challenging. Evaluation data that include information on whether evaluators have previously experienced misrecognition and on their subsequent evaluations of other candidates are often not readily available. Furthermore, causally isolating the average effect of an evaluator's experience of misrecognition on their subsequent assessments of others requires that evaluator misrecognition occurs in a random fashion and that subsequent and standardized evaluations of similar candidates are captured. To address these challenges and maintain high levels of both external and internal validity, we leverage data from three studies: A quasi-natural

experiment of investment professionals in the field and two comprehensive, pre-registered multistage survey experiments.²

STUDY 1

For an initial test of our theory, we used field data from Investor Circle (IC, a pseudonym), a private digital platform for investment professionals (professionals) to share investment recommendations with each other. To gain access to IC, a professional must show proof of current employment as an investment professional at a buy-side investment management firm (e.g., hedge fund, mutual fund). Buy-side investment professionals are responsible for researching market opportunities (e.g., stocks) with the aim of investing significant amounts of capital on behalf of their employer and its clients. Professionals on IC submit investment recommendations to buy or short sell stocks and provide a detailed justification that supports their recommendation. These justifications are comprehensive and resemble the analyses these professionals typically conduct within their firm when making investment decisions.

Once a recommendation is submitted, it is accessible to current and future members of the platform and can be viewed and evaluated by these other platform members. Professionals initially view a summary of the recommendations submitted by peers—stock name, recommended position (i.e., buy or sell), performance to date, as well as recommender’s name and employment affiliation—and once they click on a recommendation, they gain access to the details supporting that recommendation. They can also then evaluate the quality of these recommendations using a five-star (integer) rating. Thus, each professional can move between

² Pre-registrations for each survey experiment available at the following links: Study 2 https://osf.io/8tpab/?view_only=fb75fde223cf4b6e852a8bf011e034ea; and Study 3 https://osf.io/49qc2/?view_only=f9ebd111c1124b9b96dc03a295d3cdc5

being the candidate (i.e., having their own recommendations evaluated) and being the evaluator (i.e., rating the quality of recommendations submitted by others).

Quasi-Natural Experiment and Sample

Data for this study are from 2013 when a quasi-natural experiment related to misrecognition occurred on IC, allowing us to examine the relationship between underrecognition and subsequent evaluator behavior in a field setting. That year, the platform introduced a form of recognition via weekly e-mails to all users highlighting the professional and recommendation that received the highest quality rating from other professionals in the prior week. Because this recognition was granted solely based on peer ratings of quality, the process for allocating recognition in this context was meritocratic and instances of misrecognition were knowable to those who were underrecognized. Additionally, this recognition serves as a performance-based status designation common in professional contexts (Bowers and Prato 2018; Merton 1968). These professionals valued the platform as an opportunity to elicit acknowledgement from their peers.

We observed 27 instances in which two recommendations had a perfect five-star rating and were thus equally deserving of the award. In these cases, the platform employee leading the award campaign simply chose which of the two recommendations—and thus which of the two professionals—would be recognized that week. This choice was “as good as random” because the employee did not have any explicit criteria for selecting between the two equally eligible professionals. Thus, this quasi-natural experiment provides an opportunity to test an essential

element of our theory: The effect of a person experiencing an instance of underrecognition on their subsequent evaluations of others.³

We first provide evidence in support of our assumption that misrecognition occurred in a good as random way by comparing whether professionals who were correctly recognized were similar to professionals who were underrecognized. Using five observable characteristics (see Appendix A, Table A1), we find no statistical differences. However, given the small sample size, it is also important to consider substantive differences. There is no substantive difference between correctly recognized and underrecognized professionals in any measure besides *Elite Education*. We find that correctly recognized professionals are substantively more likely to have attended a highly ranked school (undergraduate or graduate) (0.59) relative to underrecognized professionals (0.44). Overall, this comparison supports the notion that professionals who received recognition were similar to those who did not.

Underrecognized professionals—as well as their counterparts who received recognition (i.e., were correctly recognized)—could provide ratings of recommendations submitted by others both before and after the award was allocated. To provide an initial test of our theory related to the effect of underrecognition, we examine whether underrecognized evaluators change their rating behavior following their experience of misrecognition. Specifically, we compare the average ratings provided to others before the award event (i.e., in the period *before* misrecognition occurred) to the average ratings after the award event (i.e., in the period *after* misrecognition occurred). By examining whether changes in rating behavior across these two

³ In this setting, we do not observe instances of overrecognition because a high rating was the criteria for identifying potential winners.

periods differed for underrecognized versus correctly recognized professionals, we can establish whether underrecognition distinctly affects evaluative behavior.

Our final analytical sample includes 952 ratings submitted by 54 unique evaluators within a 180-day window before and after the recognition event (i.e., award e-mail).

Measures

The main outcome of interest is *Rating*, which takes an integer value between 1 and 5, reflecting the rating a professional gave to recommendations submitted by others. Our main independent variable of interest captures whether the focal individual experienced misrecognition.

Professionals are classified as either *Correctly Recognized*, if they were chosen to receive recognition, or *Underrecognized*, if they were not chosen to receive recognition in a given week where there were two top-performing professionals (i.e., a tie).

Results

Professionals submitted a total of 351 ratings in the 180 days before the focal recognition (e-mail) event, and they submitted 601 ratings in the 180 days after the recognition event, representing a 70 percent increase in rating activity in the weeks following recognition events.

This increase is consistent with recent research on how status affects an individual's subsequent productivity. For example, Bol et al. (2018) found that those who won early career grants were more likely to win midcareer grants and to secure more research funding across their careers.

Importantly, this effect was driven by the fact that winners often participated *more* than nonwinners. Given this work, we examined pre-recognition and post-recognition period rating behavior among those people who were later correctly recognized (i.e., winners) and underrecognized (i.e., non-winners). In the pre-recognition period, we found no difference in rating behavior: correctly recognized professionals submitted 172 (49 percent) and

underrecognized professionals submitted 179 (51 percent) of the 351 ratings. Consistent with this prior research, however, we found that the observed increase in rating behavior in the post-recognition period was driven by correctly recognized professionals who submitted 380 of the 601 ratings (63 percent).

Next, we turned our attention to testing our main theoretical argument and Hypothesis 1a—that underrecognized evaluators will be less likely to recognize others. Specifically, we examined whether professionals who narrowly missed out on being recognized went on to give lower ratings to others by comparing the mean rating submitted by these evaluators before and after the recognition events (Figure 1A). Indeed, there is a substantial difference in the ratings submitted by underrecognized professionals across the pre- and post-recognition periods. In the pre-period, when these professionals had not yet been underrecognized, they rated others' recommendations as 2.88 stars on average. Importantly, the ratings provided in the pre-period did not significantly differ between those later underrecognized and those later correctly recognized: 2.88 versus 3.05 stars, respectively—a difference of 5.6 percent ($p = 0.255$). As evidence of the effect of misrecognition, underrecognized professionals significantly decreased their ratings in the post-recognition period, rating recommendations 17.4 percent lower than they did in the pre-recognition period (2.38 vs. 2.88; $p < 0.001$).

[FIGURE 1]

By contrast, correctly recognized professionals submitted similar ratings before and after being recognized. The average rating submitted by these professionals was 3.05 stars in the pre-recognition period and 2.96 stars in the post-recognition period, a difference of only 2.95 percent ($p = 0.439$). Therefore, ratings from underrecognized professionals represent the lowest ratings observed in this sample: nearly 20 percent lower than those from correctly recognized

professionals in the post-recognition period (2.38 vs. 2.96; $p < 0.001$). These findings are consistent when we restrict the post-recognition period to ratings submitted within 30 days following award events (Figure 1B).

This quasi-natural experiment provides field-based support for Hypothesis 1a: underrecognition affects how professionals subsequently dole out recognition to peers, such that they provide significantly lower ratings of quality. But since these data do not allow us to directly examine the effect of overrecognition and offer only limited insight into potential mechanisms, we next turn our attention to our online survey experiments.

STUDY 2

Given the constraints on identification and sample in our field data, our goal was to maximize the internal validity of our empirical approach through controlled experimental protocols and clear, effective manipulations to examine our hypotheses more fully (Campbell and Stanley 1963; Kenny 2019; Thye 2000). As such, Study 2 uses a comprehensive, pre-registered two-stage survey experiment on an online platform to replicate the effect of underrecognition found in the field (Study 1, Hypothesis 1a), extend these findings to include cases of overrecognition (Hypothesis 2a), and more directly examine the mechanism underlying the observed effect of underrecognition (Hypothesis 1b).

Materials

Our experimental design involved two sequential stages (Figure 2). In the first stage, participants completed an aptitude test for which they were evaluated and then were told whether they received our “Elite Award,” which we describe in detail below. In the second stage, participants shifted to being evaluators of another person’s responses to a similar aptitude test. Our goal was to compile a set of ten questions for the aptitude tests likely to convince participants that their

responses provide a proxy for their general ability. Consistent with prior research that has employed composite cognitive-assessment measures (Chapman et al. 2018; Dworak et al. 2021), we constructed both instruments using questions drawn from standardized aptitude-testing sources like the International Cognitive Ability Resource (ICAR-16) and MENSA IQ challenge (Vining 1984).

[FIGURE 2]

We constructed two comparable versions of the aptitude test—one to be completed by participants and another to serve as peer responses for them to evaluate—by pre-testing items for difficulty within question type and then matching them across the two versions. We grouped our initial sample of 39 questions into four types: “simple,” “challenge,” “logic,” and “pattern.” To find pairs of questions of comparable difficulty for the participant test and the peer test, we recruited 200 participants from Amazon’s Mechanical Turk (MTurk), randomly presented them 10 of the 39 questions in our set and asked them to rate the difficulty of each question on a scale of 1–10. We then used the average perceived difficulty of each question to identify matches within each question type, which produced a total of 10 pairs of equally difficult questions to be split between the two tests. For more detail on the process of constructing the aptitude tests, examples of each type of question, and results of this pre-test see Appendix B (Figures B1-B4 and Table B1-B2).

Participants

We conducted our online survey experiment using CloudResearch (Litman, Robinson, and Abberbock 2017). CloudResearch is an online labor-market platform where gig workers (MTurkers) can complete tasks and jobs, known as Human Intelligence Tasks (HITs), posted by individuals and organizations. Our research team, serving as the employer, recruited workers to

perform our task (i.e., to complete an aptitude test and evaluate a peer's responses to a comparable test) and described this task as helping the research team develop an aptitude test. MTurk has been widely used in academic research, including by sociologists (Hahl, Kim, and Zuckerman Sivan 2018; Kuwabara 2015; Leung and Koppman 2018; Miles, Charron-Chénier, and Schleifer 2019; Paxton, Velasco, and Ressler 2020; Tilcsik 2021), because it provides access to participants who exhibit the classic heuristics and biases we tend to study and who are at least as attentive to directions as participants from other samples. Using MTurk is also an economical way of running large, complex experiments (Paolacci, Chandler, and Ipeirotis 2010).

Participants

Per our pre-registration and a priori power analysis, we recruited 1,446 US-based adult participants. We received full data (i.e., all measures of interest) from 1,560 participants due to CloudResearch's tendency to oversample and high bounce rate (Litman et al. 2017). Completion of the task was estimated to take 15 minutes; participants were paid \$2.00, with a guaranteed \$0.50 bonus, which works out to an estimated hourly wage of \$10 per hour. We embedded two attention checks in the experiment (see Appendix C, Figure C2), one related to our explanation of the Elite Award and the other about whether they had received the Elite Award. Per our pre-registration, participants who answered either question incorrectly—312 participants (20 percent) in the first case and 45 (3 percent) in the second—were dropped from our analyses. Our final analytical sample consisted of data from 1,202 participants.

Procedures

Figure 2 provides an overview of the two-stage experimental design. The first stage called for participants to complete a 10-question aptitude test, and the second stage asked them to evaluate a peer's responses to a different aptitude test with 10 similar questions.

Participants were then informed that their responses to the questions on the first-stage aptitude test would be evaluated by the study team. They were also told at this time that they would be considered for the Elite Award, our high-status distinction. Our goals in describing the Elite Award were twofold: (1) to make the award desirable, so as to increase the value that participants would attribute to it (Besley and Ghatak 2005), and (2) to reveal that the award—like any prestigious distinction or valued recognition, including the recognition observed in Study 1—was in “limited supply” and intended to reward only “high performers” (Podolny 1993). Furthermore, we also designed the award to create a clear status distinction between those who were given the award and those who were not, following on recent research showing that the clarity of status hierarchies can influence material outcomes (Accominotti et al. 2022). To increase the desirability of the award, participants were told that those given the award would be given an Elite Award Certificate and early access to a future related HIT (Appendix C, Figure C1). After completing their test, participants were informed that their performance was either below or above the average of others and were told whether they had received the Elite Award. See Appendix C, Table C1 for details on the experimental language used across conditions.⁴

To increase the ecological validity of the study and to provide misrecognized participants a realistic experience, participants were given accurate information about their performance on their aptitude test—the first stage of the experiment. Our assumption was that participants would have a sense of how they performed and thus we wanted participants to receive an accurate measure of their relative performance. To achieve this, we automatically scored each participant’s responses and designated those who answered seven or more questions correctly to

⁴ Across previous iterations of this study, we varied the language used to introduce our manipulation of misrecognition, with consistent results across these iterations. See Appendix D for details on these variations and results.

be high performers while all others were designated low performers. This categorization cutoff was based on average participant performance during pilot testing and then an examination of the first set of responses in the experiment (see Appendix B: Table B2, Panel A, for the average percentage of correct responses by question, and Table B2, Panel B, for the distribution of the total number of questions answered correctly). Participants were simply told whether they had answered more (high performer) or fewer (low performer) questions correctly than average.

Once a participant's test was scored, we randomly assigned participants to either receive or not receive the Elite Award. Participants were thus assigned to one of four conditions: (1) high performers who received the Elite Award (correctly recognized high performer), (2) high performers who did not receive the Elite Award (underrecognized), (3) low performers who did not receive the Elite Award (correctly recognized low performer), and (4) low performers who received the Elite Award (overrecognized).

During the second stage of the task, participants were told that they would evaluate a peer test submitted by someone else; however, these responses were devised by the researchers and held constant across conditions. All participants were presented with an identical peer test where five questions were answered correctly and five were answered incorrectly. Participants were then tasked with evaluating whether each of the 10 responses was correct (see Appendix B, Figure B5, for a sample of the peer responses presented to participants). Our focus is on comparing the propensity for evaluators to grant the Elite Award based on their perception of how the peer performed. Namely, participants were neither aware of how many of these responses were indeed correct or incorrect nor informed about the number of correct responses necessary to be deemed a high performer. After evaluating each of the peer's responses,

participants were asked about their overall perception of the peer's performance and whether they recommended that the peer be recognized with the Elite Award.

Measures of Interest

To capture evaluators' overall perception of how their peer performed, we asked: "Overall, how well do you think this participant performed relative to others who have taken this test?"

Perceived Peer Performance is based on responses on a 1–7 scale: from (1) very poorly to (7) exceptional, with a midpoint of (4) good. Thus, higher values on the scale correspond to an evaluator perceiving their peer to be a stronger performer. We then asked evaluators whether they recommended that the peer be recognized with the Elite Award. *Recommend Elite Award*, which serves as our main dependent variable, takes the value of 1 if the participant answered "yes" and 0 if "no."

We also conducted analyses with control variables. These include self-reported measures and data collected through the experiment. Participants self-reported their age (*Age*), education (*College Educated*), familiarity with cognitive assessments (*Experience Cognitive Assessments*), gender (*Female*), political orientation (*Democrat* and *Republican*), race (*White*), and work status (*Full-Time Employee*). We also collected two sets of measures capturing participants' behavior throughout the experiment. The first set pertained to the aptitude test completed by the participant in the first stage of the experiment: *Participant Test Time Spent* is the number of seconds that a participant spent completing the aptitude test; and *Participant Test Number Correct* is the number of questions the participant answered correctly. The second set pertained to the participant's evaluation of the peer test: *Time Spent on Peer Test* measures the amount of time the participant spent evaluating the peer's responses; *Number Marked Correct on Peer Test* is the number of the peer's responses to questions that the participant marked as correct; and

Accuracy on Peer Test is the accuracy of the participant's evaluation of the peer's responses. Accuracy was calculated as the percentage of the peer's responses to questions participants marked accurately as correct or incorrect.

Results

As is typical in experimental research, we sought to ensure that we effectively simulated an experience of misrecognition among participants randomly assigned to our underrecognition and overrecognition conditions (i.e., manipulation check). Specifically, we asked participants the extent to which they considered their Elite Award outcomes correct: “[Not] Being recognized with the Elite Award for my performance on the test felt correct,” measured on a 1–7 scale. As expected, correctly recognized participants were more apt to agree that their recognition was correct (5.0 for high performers and 4.7 for low performers; average 4.8) than those who were not correctly recognized (3.2 for underrecognized and 3.3 for overrecognized; average 3.2) with this difference between correctly recognized and misrecognized participants being statistically significant ($p < 0.001$).

Descriptive Statistics

Table 1 provides descriptive statistics and Appendix C, Table C2 provides a correlation matrix for our main variables. On their own aptitude tests, participants answered an average of 6.7 of the 10 questions correctly. In their role as evaluators, participants reported that their peer had answered 4.8 questions correctly, on average, which is in line with the actual number of correct responses on the peer test (5). On average, participants accurately identified peer responses as correct or incorrect 76 percent of the time. Evaluators perceived the peer's overall performance as 3.9 (good) out of 7.0, on average, and recommended that the peer receive the Elite Award 38 percent of the time.

Demographically, participants were on average, 44 years old with 57 percent self-reporting as female, 74 percent as White, 57 percent as college-educated, 47 percent as Democrats, 20 percent as Republicans, and 58 percent as full-time employees. A randomization table for all variables can be found in Appendix C, Table C3.

[TABLE 1]

Effect of Recognition on Evaluator's Propensity to Recommend Elite Award to Others

To test our first hypotheses—that underrecognized (overrecognized) evaluators will be less (more) likely to grant recognition to others—we turn our attention to whether an evaluator's propensity of granting the Elite Award to a peer varies across conditions. Table 2 (Model 1) shows the effect of our main conditions with controls and Figure 3 provides the marginal effects from this regression. Our focal comparison is whether the propensity to grant recognition (i.e., recommend the Elite Award) varies for misrecognized participants relative to correctly recognized counterparts who performed similarly.

These results provide causal evidence that misrecognition is reproduced. Among high performers, correctly recognized evaluators—those who received the Elite Award—granted the Elite Award to a peer 44 percent of the time; meanwhile, underrecognized evaluators—those who did not receive the Elite Award—granted the award only 27 percent of the time, or about 39 percent less often ($p < 0.001$). Thus, consistent with Study 1 and in support of Hypothesis 1a, we find that underrecognition decreases the likelihood that evaluators grant recognition to others.

Among low performers, correctly recognized evaluators—those who did not receive the Elite Award—granted recognition 26 percent of the time. In sharp contrast, overrecognized evaluators—low performers who nonetheless received the Elite Award—granted recognition to a peer 59 percent of the time, more than twice as often as their correctly recognized counterparts

(127 percent more often; $p < 0.001$). Thus, in support of Hypothesis 2a, we find that overrecognition significantly increases the likelihood of evaluators to grant recognition to others.

[FIGURE 3]

[TABLE 2]

Overall, these results provide causal evidence that experiencing misrecognition alters evaluators' likelihood to subsequently grant recognition to others. Consistent with Study 1, underrecognized evaluators were significantly less likely to grant recognition to others than similar high-performing evaluators who (correctly) received the award themselves. Expanding on our results from Study 1, Study 2 also provided insight into the causal effect of overrecognition (H2a): Lower-performing evaluators who were nonetheless granted the Elite Award were more than twice as likely to grant recognition to others than similar performing evaluators who (correctly) did *not* receive the status award. Importantly, our design ensured that all participants across recognition conditions evaluated an identical quality peer. As such, these differences can be fully attributed to misrecognition. Furthermore, these results are robust even when including those who failed attention checks (Appendix C: Table C4 Model 1, Figure C3). Taken together, these results support our first set of hypotheses (H1a and H2a).

Since we provided accurate information to participants regarding their relative performance on the aptitude test, we can further scrutinize the generalizability of our results by examining whether the effect of misrecognition is localized among participants who had the most pronounced misrecognition experience: That is, whether our effects are driven by those who were among the very top (or very bottom) performers and did not receive (or did receive) the Elite Award, and thus may have had the most pronounced misrecognition experience. Figure 4 plots the relationship between our four recognition conditions and the propensity to grant the

Elite Award by the number of questions the participant answered correctly.⁵ These results reveal the robustness of our findings since the effect of underrecognition and overrecognition emerges at all performance levels. Regardless of the number of questions participants answered correctly, underrecognized evaluators were—relative to correctly recognized evaluators—significantly *less* likely to recommend the award while overrecognized evaluators were significantly *more* likely to do so.

[FIGURE 4]

It is worth highlighting the comparison of the likelihood to recommend the Elite Award by those evaluators at the cutoff of being labeled either a low or high performer (i.e., those participants who answered six versus seven questions correctly). These analyses mirror Merton's example of the 41st chair (1968). This subset of similar participants demonstrated nearly identical ability, but a one question difference in the number answered correctly dictated whether they were classified as low or high performers. The propensity for evaluators to grant recognition among this subset replicates our main results, providing the clearest evidence of the causal effect of being misrecognized on a person's subsequent evaluations of others.

Is Misrecognition Being Reproduced?

The fact that underrecognized participants grant recognition at a lower rate—and overrecognized participants at a higher rate—is consistent with our theoretical argument that evaluators who experience misrecognition are more apt to subsequently produce unmeritocratic evaluative outcomes. However, to more directly examine the extent to which such reproduction emerges, it is useful to consider whether underrecognized evaluators are less likely to recognize *even those*

⁵ We grouped together participants who answered three or fewer questions correctly and participants who answered 9 or 10 questions correctly because they represent less than 5 percent of our sample. One participant was put in the incorrect condition and was removed from this analysis.

they perceive to be top performers. Similarly, overrecognized evaluators granting recognition at a higher rate *even to those they perceive to be poor performers* would provide clear evidence of the reproduction of overrecognition.

To make this comparison, we relied on our pre-registered measure of an evaluator's overall opinion of their peer's performance (*Perceived Peer Performance*). Given the evaluative criteria we outlined stated that the Elite Award be granted to high-performing candidates, those perceived as stronger performers should be the ones to be recognized with the award. Therefore, we focus on comparing the likelihood for evaluators to grant recognition across conditions at different levels of perceived peer performance. Figure 5 plots the marginal effects of a regression predicting *Recommend Elite Award* by each condition interacted with the evaluator's *Perceived Peer Performance* (Appendix C, Table C5). Given the rarity of *Perceived Peer Performance* scores at the lowest and highest levels and to facilitate interpretation of our effects, we grouped the scores into three buckets: perception that the peer was a weak performer (1-3), a neutral performer (4), or a strong performer (5-7). Consistent with the general notion that recognition—in this case the Elite Award—is reserved for high performers, *Perceived Peer Performance* had a positive average effect on an evaluator's likelihood of recommending an Elite Award for a peer (Table 2 Model 1). However, the relationship between perceived performance and recommending the Elite Award varied across conditions in meaningful ways.

First, these results provide evidence that the lower likelihood for underrecognized evaluators to grant recognition to others in fact reflects a tendency for these evaluators to reproduce underrecognition in their subsequent evaluations. Among high performers, when *Perceived Peer Performance* was highest (5 to 7), correctly recognized evaluators granted the award 77 percent of the time while underrecognized evaluators granted it 51 percent of time ($p <$

0.001). The fact that underrecognized evaluators were 34 percent *less* likely to recognize those whom they assessed to be the highest performers provides clear evidence that underrecognized evaluators reproduce their experience of misrecognition: They grant recognition far less than their correctly recognized counterparts even to those they deem high performers.

Meanwhile, when *Perceived Peer Performance* was lowest (1 to 3), correctly recognized low performing evaluators granted recognition only 2 percent of the time whereas overrecognized evaluators granted recognition 38 percent of the time ($p < 0.001$). The fact that overrecognized evaluators were 19 times *more* likely than correctly recognized low performers to recognize those whom they assessed to be low performers provides further evidence that overrecognized evaluators also reproduce their experience of misrecognition by overrecognizing those they deem to be poor performers. Taken together, these analyses provide consistent support for Hypotheses 1a and 2a, showing that experiences of misrecognition lead evaluators to misrecognize subsequent others.

[FIGURE 5]

Examining Fairness as a Driver of the Effect of Underrecognized Evaluator Behavior

We theorized that misrecognized evaluators are apt to see the process for allocating recognition as fundamentally unfair, or at least as less equitable, than do their correctly recognized counterparts. In addition to violating our specification that the allocation of the award was reserved for top performances on the aptitude test, this experience also contradicts general meritocratic beliefs that rewards ought to be granted only to the best candidates (and not to weaker performers). Thus, high performers who are denied recognition and low performers who are nonetheless recognized are likely to perceive the evaluation process as relatively less fair.

That said, we built on theories of positive and negative inequity to hypothesize that these perceptions of unfairness would drive the subsequent evaluative behavior of underrecognized but not of overrecognized evaluators (Hypothesis 1b). We test this hypothesis using our pre-registered measure of *Fairness Receiving Award*. After participants evaluated their peer, we asked them to score on a 1-7 scale the extent to which they agreed with the statement: “The process for determining whether participants are recognized with the Elite Award felt fair.” Higher values indicated that the participant saw the process as fairer whereas lower values represent perceptions that the process was less fair. Consistent with our argument, misrecognized evaluators viewed the process of allocating the Elite Award as less fair than did correctly recognized evaluators (average of 3.51 vs. 4.75; $p < 0.001$).

We dichotomized our *Fairness Receiving Award* variable: *Unfairness* takes a value of 1 for those indicating the process was relatively less fair (responses from 1-3) and equal to 0 for those responding that the process was neutral to fairer (responses from 4-7). Examining the effect of *Unfairness* across conditions allows us to determine whether perceptions of (un)fairness uniquely drive underrecognized evaluators’ subsequent recognition behavior, as we predicted.

As we hypothesized, we find that perceptions of unfairness only drive the propensity to recommend the Elite Award for underrecognized evaluators. Specifically, we find that underrecognized evaluators grant recognition approximately 15 percent of the time when they see the process as unfair; however, when they perceive it to be fair, they are 73 percent more likely to grant recognition to others (26 percent of time; $p = 0.008$). The likelihood of giving the award as a function of *Unfairness* did not differ in any of the other three conditions (see Figure 6 for marginal plots).⁶ These results are substantively unchanged when we include participants

⁶ The non-significant main effect of *Unfairness* in Table 2 is because the reference group is the “correctly recognized, high performer” condition. This effect indicates that unfairness does not affect propensity to grant the

who failed either attention check (Appendix C: Table C4 Model 2, Figure C4). These findings support our theory and Hypothesis 1b that perceived (un)fairness is a key driver of underrecognized—but not overrecognized—evaluators’ subsequent behaviors because experiences of negative inequity loom larger than experiences of positive inequity. Thus, perceptions of unfairness among underrecognized (but not overrecognized) evaluators meaningfully influence their likelihood to give the award to others.

[FIGURE 6]

STUDY 3

Study 1 and Study 2 provide convergent and convincing evidence of our theory and support for Hypotheses 1a, 1b, and 2a. Thus, Study 3 focuses on testing Hypothesis 2b by examining whether informational cues drive our observed pattern of results for overrecognized evaluators.

Study 3 uses the same survey design as Study 2, but some participants are randomly assigned to receive an evaluative criteria manipulation *before* they assess whether the peer should receive the award (see Figure 7). Participants assigned to receive this additional criteria manipulation are reminded that the award should go only to top performers and are given additional information about the performance threshold necessary for granting the award. To the extent that informational cues are driving overrecognized evaluators to overrecognize others, adding this detailed information about how one ought to grant recognition should attenuate this effect, effectually reducing overrecognized evaluators’ elevated propensity to subsequently grant recognition. Specifically, if overrecognized evaluators are making inferences about how to

award among our reference group (correctly recognized high performers). However, the relevant comparison for testing our hypothesis that perceptions of unfairness underlie the effect of misrecognition—particularly among underrecognized evaluators—is between underrecognized participants who believed the allocation of the Elite award was fair versus underrecognized participants who perceived it to be unfair, which is not easily discernable from the model and thus presented in Figure 6.

evaluate others from their experience having been overrecognized, as we hypothesize (H2b), they ought to grant recognition more similarly to their correctly recognized counterparts when they are given explicit information on how to evaluate (i.e., information that is more relevant than the cues available through their experience of overrecognition).

[FIGURE 7]

Participants and Procedures

Per our pre-registration and a priori power analysis, we recruited 2,140 US-based adult participants for our 2 (High Performance, Low Performance) X 2 (Elite Award, No Elite Award) X 2 (Additional Criteria, No Additional Criteria) between-subjects design (see Figure 7). We received full data from 2,320 participants due to CloudResearch's tendency to oversample. Completion of the task was estimated to take 15-20 minutes; participants were paid \$3 for an estimated hourly wage of about \$10 per hour. See Appendix D, Table D1 for full descriptives, Table D2 for randomization tables, and Table D3 for a correlation matrix of all variables.

During the study, participants were asked whether they would recommend the Elite Award, using nearly the same language as in Study 2. Participants in the additional criteria condition were told: "Next you will be making a recommendation about whether this person should get the Elite Award. As a reminder, this award should be given to the high-performing participants. Typically, participants who get at least 7 correct get the award. *Would you recommend this participant be recognized with the Elite Award?*" Those in the no additional criteria conditions, however, were simply asked "*Would you recommend this participant be recognized with the Elite Award?*" (the same as Study 2).

Our goal with adding the additional explicit criteria was to examine whether overrecognized evaluators were using procedural informational cues from their experience of

misrecognition as guideposts on how to recognize others. Stated otherwise, to what extent were these evaluators inferring how to allocate recognition from their prior experience having been evaluated? Since the additional criteria reduces ambiguity about how an evaluator ought to evaluate others, we would expect overrecognized participants in the additional criteria condition to grant recognition at a *lower* rate than those in the no additional criteria condition. Thus, receiving the additional criteria ought to lead overrecognized evaluators to grant recognition more similarly to their correctly recognized counterparts. Conversely—although we do not theorize that this will occur—if informational cues also underlie the effect of underrecognition, we might expect that underrecognized evaluators in our additional criteria condition would grant recognition at a *higher* rate than those in the no additional criteria condition. In other words, to the extent that informational cues drive misrecognized evaluators’ behaviors, our additional evaluative criteria manipulation should attenuate the reproduction of misrecognition, such that misrecognized evaluators grant recognition more similarly to their correctly recognized counterparts.

We used the same attention checks as in Study 2, but also added a check following the additional criteria intervention to ensure people were attentive to this treatment about the criteria for granting the award. Consistent with our pre-registration, participants who answered any of the three attention checks incorrectly—the benefits of the Elite Award (489 participants; 21 percent), whether or not they received the Elite Award (60 participants; 3 percent), or who should receive the award (31 participants; 1 percent)—were dropped from our analyses. Our final analytical sample consisted of data from 1,740 participants.

Results

We first replicated our pattern of main results from Study 2 (Figure 3) for participants who did not receive additional criteria; these results are presented in Appendix D, Table D4, Model 1. Among high performers, underrecognized evaluators were approximately 21 percent less likely to grant recognition to others than correctly recognized evaluators ($p = 0.045$). Also consistent with Study 2, among low performers, overrecognized evaluators were 124 percent more likely to recognize others than were correctly recognized evaluators ($p < 0.001$). The results are robust to the inclusion of those who failed any of the three attention checks (Appendix D: Figure D2 and Table D4, Model 2).

Informational Cues as a Driver of the Effect of Overrecognized Evaluator Behavior

We next turn our attention to the unique intervention in Study 3 aimed at assessing whether overrecognized evaluators' behaviors were driven by informational cues, as we hypothesized. Specifically, we examined whether providing additional information about the appropriate criteria shaped overrecognized evaluator recognition behavior. We thus examined variation in the likelihood to grant the Elite Award based on whether the evaluator received the additional criteria or not, where our comparisons focused within recognition condition (e.g., whether the additional criteria changed evaluative behavior among overrecognized participants; Figure 8A).

[TABLE 3]

In support of Hypothesis 2b, we find evidence that informational cues underlie the effect of overrecognition. For correctly recognized, low performers, the additional evaluative criteria had no effect on their recognition behavior: Those who did not receive the additional criteria language gave the award at similar rates to those who did receive the criteria language (27 vs. 30 percent, $p = 0.463$). Among overrecognized evaluators, however, the additional criteria language

significantly reduced their propensity to grant recognition to others. We find that those who did not receive the additional criteria manipulation gave the award about 64 percent of the time while those who received the criteria gave the award about 53 percent of the time (a 17 percent decrease; $p = 0.017$).

Moreover, and consistent with our theorizing, we do not find evidence that informational cues strongly underlie underrecognized evaluators' behaviors. Among high performers who received the award (i.e., correctly recognized, high performers), providing them with the additional criteria had no effect on their likelihood of subsequently granting recognition to others (34 vs. 34 percent, $p = 0.865$). Similarly, this additional criteria had only a marginally significant effect on underrecognized evaluators' likelihood of granting recognition (27 vs. 22 percent, $p = 0.088$). Thus, informational cues meaningfully drive the evaluative behavior of overrecognized evaluators: Telling them *how* they should evaluate reduces their propensity to overrecognize others.

As further evidence that informational cues uniquely drive evaluative behavior for overrecognized evaluators, we find the additional criteria has the strongest effect when we limit our analyses to participants who marked fewer than 7 questions correct on the peer test—a performance level that is below that stated as requisite for granting the award in the additional criteria manipulation (Figure 8B). Specifically, among overrecognized evaluators who indicated the peer answered 6 or fewer correct (below this stated cutoff), the additional criteria significantly reduced the reproduction of overrecognition: Whereas overrecognized evaluators who did not receive the additional criteria granted recognition 62 percent of the time, those who received the additional criteria granted recognition in 46 percent of these cases (a 16 percent decrease; $p = 0.001$).

[FIGURE 8]

This study thus provides causal evidence that our proposed informational cues mechanism primarily underlies overrecognized evaluator behavior. In other words, overrecognized evaluators reproduce overrecognition because they rely on insights from their own experience for determining how to evaluate others. Reminding overrecognized evaluators of the award criteria—and, by extension, how to effectively grant recognition to others—materially attenuates the tendency for overrecognition to be reproduced. Importantly, and critical to our theorizing, we find that our intervention does not meaningfully shift evaluative behavior for those who have been *underrecognized* or those who have been *correctly recognized*.

FURTHER EXPLORING GENERALIZABILITY: TWO ADDITIONAL STUDIES

Although we have observed a robust effect of misrecognition on the subsequent allocation of recognition both in the field and across our two experiments, we wanted to further examine the generalizability of our theory and results.⁷ Across our studies, experiences of misrecognition were relatively easy to discern because individuals had some information about their performance and clarity on the criteria for granting recognition. Considering evaluative contexts more generally, however, there is variance in the degree of clarity people have about their own performance and about the prevailing criteria used to grant recognition. Therefore, we conducted two additional experiments aimed at establishing generalizability in contexts where there was less clarity on each of these dimensions.

⁷ The pre-registrations for Additional Study A can be found here: https://osf.io/827fh?view_only=4c182d5e5c6e4e199026fbc6434cb970; the pre-registration for Additional Study B can be found here: https://osf.io/va3jy?view_only=89fa89a0aea84b6786b311ab0788d75f

Additional Study A: Clarity about Performance

In each of our prior studies, participants had a relatively clear sense of how well they performed. This was especially true in Study 1, but even in our survey experiments participants were informed whether they performed above or below average. By providing them with some concrete information about their performance, instances of misrecognition were relatively detectable: By knowing their performance, people can more easily detect instances where their own outcome did not align with their performance. Although individuals are given explicit performance information about their performance across a variety of contexts, there are also cases where this information is not explicitly available, leaving individuals to infer how they performed from their own self-assessment or vague performance feedback. Thus, the goal of this study was to test whether our main effect—that misrecognition influences subsequent recognition behavior—generalizes to cases where individuals do not receive concrete performance information and are instead left to make inferences.

We recruited 1,446 US-based adult participants from MTurk. We received full data from 1,463 participants. This study followed the exact same experimental design as Study 2 with one important omission: Participants did not receive information about their performance relative to others. Instead, after completing their aptitude test, they were simply notified whether they received the Elite Award or not. Thus, we still maintain our four conditions and used the same attention checks as in Study 2. In keeping with the terms of pre-registration, participants who answered either question incorrectly—328 participants (22 percent) in the first case, 27 participants (2 percent) in the second case—were dropped from our analyses. Our final analytical sample consisted of data from 1,108 participants.

Given that in this study we withheld performance information, we wanted to ensure participants still detected whether they were correctly recognized or not, as in the other studies. Our results revealed that participants in the misrecognized conditions (i.e., underrecognized and overrecognized conditions) did indeed feel a sense of misrecognition; specifically, we find that those who were correctly recognized felt their evaluative outcome was correct more than those who were not correctly recognized (4.79 vs. 4.19, $p < 0.001$). In a way this result reaffirms the prevalence of feeling underrecognized and overrecognized—people are apt to sense when they are misrecognized even if their exact performance result is not communicated.

Our focus here is comparing the likelihood for evaluators to grant recognition to others across our conditions. We find that among high performers, correctly recognized evaluators recommended that a peer be recognized with the award 45 percent of the time, whereas underrecognized evaluators recommended it only 29 percent of the time ($p < 0.001$). Underrecognized evaluators were thus approximately 60 percent less likely to grant recognition to others, replicating our main effects and consistent with Hypothesis 1a.

Among low performers, overrecognized evaluators were 113 percent more likely to recognize others in their subsequent evaluations than were correctly recognized evaluators ($p < 0.001$). Specifically, whereas correctly recognized low performers recognized others about 23 percent of the time, overrecognized evaluators did so about 49 percent of the time. Again, these results replicate the main results from the previous studies and are consistent with Hypothesis 2a. Thus, we find that even when evaluators do not have direct information about how they performed, misrecognition still influences their propensity to grant recognition to others. This finding thus provides further ecological validity of our theory and experimental approach.

Additional Study B: Clarity of Evaluative Criteria

In our main studies (Study 2 and Study 3), we also told participants that the award should be given to the “high” or “best” performers. Given that such clarity about evaluative criteria are not available to evaluators in all evaluative processes, we intentionally made the evaluative criteria vague in this additional study. Leveraging a similar experimental design as Study 2, participants were told that the Elite Award would be given to those who we believed would “*add the most value*” instead of those who were “*high performers*.” This modification naturally introduces a greater level of ambiguity about how recognition is granted and thus is apt to make it more difficult for participants to discern that they have experienced misrecognition. As such, finding that misrecognition is nonetheless still reproduced would serve to strengthen the generalizability of our theory and findings.

We recruited 2,480 US-based adult participants from MTurk, 2,472 of whom finished the entire task. This study followed the exact same experimental design as prior studies with the slight change mentioned above: Participants were simply told the award would be given to those who “add the most value,” without explicating more precise criteria or what constitutes adding value. Using the same two attention checks as Study 2 and Additional Study A, we dropped all participants who answered either question incorrectly—474 participants (19 percent) in the first case and 32 (1 percent) in the second. Our final analytical sample consisted of data from 1,966 participants.

We find that the results from Study 2 are replicated even when the evaluative criteria were less clear, such that misrecognition is reproduced. Among high performers, correctly recognized evaluators—those who received the Elite Award—granted recognition to others 39 percent of the time; underrecognized evaluators—those who did not receive the Elite Award—

granted recognition only 13 percent of the time, or about 66 percent less often ($p < 0.001$). Among low performers, correctly recognized evaluators—those who did not receive the Elite Award—granted recognition about 33 percent of the time; overrecognized evaluators—those who received the Elite Award—granted recognition about 75 percent of the time, or over twice as often ($p < 0.001$). Thus, even in cases where evaluative criteria are vague, people still experience instances of misrecognition *and* reproduce those experiences in their subsequent assessments of others.

CONCLUSIONS AND DISCUSSION

Misrecognition is common even in evaluation processes that strive to identify and recognize only the best candidates (Botelho and Abraham 2017; Bowers and Prato 2018; Castilla 2008; Dencker 2009; Lynn et al. 2009; Paik et al. 2023). Implications for those experiencing such unmeritocratic evaluative outcomes are well established, yet existing research has not examined whether there are broader downstream effects of misrecognition on subsequent evaluations. Quite often, individuals move back and forth from being evaluated to serving as evaluators (e.g., in peer evaluation systems Brett and Atwater 2001; Siler et al. 2015). In this paper, we thus shift the focus to identifying the downstream effects of misrecognition for how people later evaluate others within the same evaluation system. Specifically, we examine whether misrecognition is reproduced in these subsequent evaluations.

We theorized that previously misrecognized evaluators would be more apt to deviate from meritocratic ideals prescribing that recognition only go to top performers, even in systems explicitly aimed at fulfilling this objective. Across a quasi-natural experiment in the field and two online survey experiments, we test our theory and show that misrecognized evaluators—those who do not receive the recognition or rewards expected for their level of performance—

differ in how they subsequently evaluate their peers compared to those who performed similarly (whether high or low) and were correctly recognized. Whereas those who previously experienced underrecognition were less likely to grant recognition to others, those who experienced overrecognition were significantly more apt to allocate recognition subsequently. Importantly, our results also reveal that the underlying drivers leading to this reproduction of misrecognition are distinct and a function of the type of misrecognition experienced. Underrecognized evaluators are far less likely to grant recognition to others due to perceptions that they had been evaluated unfairly whereas overrecognized evaluators grant recognition to others more freely because of the informational cues they access through their experiences of having been overrecognized.

First—and most centrally—these results contribute to sociological research on evaluations. Although research on meritocracy is often situated within an evaluative context, we directly bridge these areas of research to uncover the downstream effects of experiencing unmeritocratic evaluative outcomes for how people subsequently evaluate others. Scholars have recently begun to consider how an evaluator’s prior experiences being evaluated affect the way they later evaluate and allocate recognition to others. Specifically, this research has focused on the valence of one’s prior evaluations, suggesting that both positive and negative prior experiences can lead people to be more objective and meritocratic (e.g., Botelho and Gertsberg 2021; Castilla and Ranganathan 2020). One conclusion from this burgeoning research is that *any* prior experience being evaluated/recognized may strengthen evaluators’ subsequent adherence to norms of meritocracy.

We build on this line of inquiry by integrating broader sociological and psychological theories related to equity, procedural fairness, and role-fulfillment to uncover an important

boundary condition: Prior experiences of being misrecognized during an evaluation *do not* lead those people to be more meritocratic in subsequent evaluator roles. Instead, underrecognized and overrecognized evaluators may be most apt to deviate from meritocratic ideals of reserving recognition only for top-performing candidates. Importantly, this deviation manifests in divergent ways for the misrecognized, such that misrecognition tends to be reproduced in kind. By highlighting that evaluative outcomes are not simply positive or negative but vary in terms of whether they are—or at least, are perceived to be—meritocratic, we thus develop a more comprehensive understanding about the link between evaluators’ prior experiences and their subsequent evaluative behavior. Our results show that the degree to which the recognition an evaluator has received aligns with their performance—and not just the valence of that prior experience—directly affects how they later evaluate others. Documenting how a person’s prior experiences with misrecognition affects their evaluative behavior is crucial for understanding the conditions under which evaluation processes fail in accurately recognizing top performers.

Second, this research contributes to a growing body of sociological research on the implications of the design and structure of evaluative systems for producing objective and unbiased outcomes (Abraham, Botelho, and Lamont-Dobbin 2024; Botelho and Abraham 2017; Botelho and Gertsberg 2021; Correll et al. 2020; Lucas et al. 2021; Rivera and Tilcsik 2019). A critical and related implication of these findings is that initial unmeritocratic outcomes become reproduced in evaluation systems where misrecognized candidates later serve as evaluators. In the case of evaluation systems aimed at allocating awards or recognition, the value of the recognition is often rooted in the assumption that a given accolade is strongly correlated with a candidate’s underlying quality. In other words, receiving the recognition is a signal that one is a top performer (Lynn et al. 2009; Podolny 1993; Simcoe and Waguespack 2010).

By considering the downstream effects of initial failures in recognizing top candidates (i.e., through instances of misrecognition), this paper suggests that evaluative systems risk increasing the likelihood of unmeritocratic outcomes when misrecognized evaluators move from being candidates to being evaluators. For instance, overrecognized evaluators grant recognition at a higher rate, even to those whom they see as the weakest performers. A strong interpretation of our results would suggest that these initial instances of misrecognition will be *consistently reproduced*, simply leading to the breakdown of evaluative processes that rely on misrecognized evaluators. However, we caution against such an interpretation and highlight two important factors to consider. First, it is important to note that misrecognition is not the modal outcome—these systems generally correctly identify candidates. Second, and even more importantly, our results show that when misrecognition does happen, there is a higher likelihood that it will be reproduced by misrecognized evaluators. That said, we do not find that misrecognized evaluators uniformly misrecognize others. Our research uncovers an additional pathway through which these systems can be imperfect: Initial instances of misrecognition have the potential to meaningfully erode the link between candidate quality and the allocation of recognition in evaluative systems.

Third, our findings provide evidence about the two distinct mechanisms behind the tendency of misrecognized evaluators to reproduce their prior experiences in subsequent evaluations of others. As we showed in Study 2, both underrecognized and overrecognized evaluators accurately perceived their outcomes as similarly unfair. However, this unfairness only motivated underrecognized evaluators to differentially recognize subsequent peers. By contrast, perceptions of unfairness did not play a substantive role in driving behavior for overrecognized evaluators. These unilateral effects of unfairness follow from our theoretical arguments that

people tend to be most reactive to negative inequity (Austin and Walster 1974; Scheer et al. 2003), such as that experienced by underrecognized evaluators.

Given that our first two studies did not allow for a direct examination of informational cues—our proposed driver of overrecognized evaluators' subsequent behavior—we designed Study 3 to achieve this aim. And indeed, we find that the propensity of overrecognized evaluators to grant recognition at a higher rate is largely driven by the inferences they draw from their experiences of misrecognition about how they ought to evaluate others. One explanation for why overrecognized, but not underrecognized, evaluators take cues from their prior experiences relates to the notion that positive inequity leads people to experience tension (e.g., guilt, imposter syndrome), which they often strive to redress (Adams 1965; Carrell and Dittrich 1978). One way they do so is by exerting more effort and improving performance (Greenberg 1988; Pritchard, Dunnette, and Gorgenson 1972). Therefore, unlike underrecognized evaluators, those who were previously overrecognized may be particularly attentive to available informational cues on how they are expected to evaluate others; in other words, they will be most focused on figuring out the appropriate way to evaluate others. It is important to note that though misrecognition may shape evaluator behavior consciously—such that underrecognized and overrecognized evaluators intentionally misrecognize others—the effects of misrecognition may also be implicit. Isolating the precise psychological motives driving misrecognized evaluators' behaviors is an opportunity for future research to further advance theory on the effects of misrecognition on subsequent evaluator behavior.

In this way, our work highlights two important levers for both researchers and policymakers to keep in mind when considering how to potentially curb the problematic effects of misrecognition. First, we found that it is possible to attenuate the reproduction of

overrecognition by providing overrecognized evaluators with clarity on how to satisfy their role through clear information about the evaluation process. In doing so, we were able to correct the behavior of these evaluators. Importantly, we did not find this approach meaningfully remedies the reproduction of underrecognition. Our research thus suggests a structural solution for attenuating the reproduction of overrecognition in evaluation systems where individuals move between being candidates and evaluators. It seems imperative to implement formalized and explicit evaluative criteria to lower the likelihood that overrecognition is reproduced. These points are consistent with the body of research on the advantages of formalized and structured evaluation processes for promoting consistency in how evaluators grant recognition (Castilla 2015; Castilla and Benard 2010). Second, our findings suggest that increasing perceptions of procedural fairness may mitigate the effects of underrecognition. Future research is needed to uncover how decision makers can effectively boost perceptions of fairness for those who feel they should have been recognized (e.g., communicating how it happens or what redress they can seek).

A key insight from our work is how readily prior experiences of misrecognition shape evaluator behavior. Across all three studies, our results show that singular experiences with misrecognition are enough to markedly shape subsequent evaluative behavior, on average. In fact, most instances of misrecognition are similarly singular: A top-performing scientist who does not receive an expected award experiences an instance of underrecognition and an employee who receives an unexpected promotion is apt to experience an instance of overrecognition. Importantly, any individual experiencing an instance of misrecognition—both in our studies and more generally—has inevitably had a *range of other prior experiences* that may or may not align with the focal misrecognition experience. Whereas some people may have

had relatively consistent experiences with either overrecognition or underrecognition, it is more likely that people have experienced a mix of each type of misrecognition and instances of correct recognition. Because the focal misrecognition (or correct recognition) experience occurred at random in our studies, our results reveal the average effect instances of misrecognition have on people's subsequent evaluative behavior in the same evaluation system. In other words, we show that across people with a range of past experiences, misrecognition is reproduced. This is consistent with research on the rapid emergence of status beliefs (Ridgeway and Correll 2006): We observe that, on average, a single experience of misrecognition powerfully shapes an evaluator's cognition and becomes more relevant than foundational norms of meritocracy or the evaluation system's underlying goal to recognize only top performers.

Though directly exploring such variation is out of scope for the current research, it is plausible that the nature of an evaluator's prior experiences with misrecognition meaningfully alters the extent to which they reproduce misrecognition. On the one hand, experiences of misrecognition may be additive, such that those who have had more consistent prior experiences with misrecognition would be most apt to reproduce misrecognition in their evaluations of others. In fact, having repeat experiences of misrecognition would arguably intensify the salience of both the available information on how to grant recognition *and* that person's perceptions of unfairness. However, we caution against drawing this conclusion and expect the relationship to be more complex. For instance, when we consider systemic inequalities and biases, it is likely that historically marginalized groups (e.g., racial minorities, women, those with lower socio-economic status) are more apt to have had repeat experiences of underrecognition. Yet existing research suggests that members of marginalized groups tend to advocate for instances where others are treated unfairly (Cohen and Huffman 2007; Greenberg and Mollick 2017; Saguy et al.

2020). Therefore, advancing this line of inquiry necessitates future research to carefully consider how repeat and consistent experiences of misrecognition shape subsequent evaluative behavior in distinct ways.

Relatedly, we are limited in what we can say about the persistence of the effects of misrecognition over time, where people likely have additional experiences being evaluated. Our field study suggests the effects do persist; examining the 180-day period after the initial misrecognition in Study 1—whereby underrecognized evaluators have inevitably been evaluated again and thus experienced varying degrees of subsequent outcomes (i.e., over/under/correct recognition)—indicates that the effect of underrecognition lasts beyond an immediate period. In our online survey experiments, however, the misrecognized evaluators were tasked with assessing a candidate soon after experiencing misrecognition. Given more time, individuals might either rationalize their misrecognition as warranted or become more convinced that it was not, leading to either weakening or strengthening of our key findings. Future research can thus advance this line of work by examining whether the effects of misrecognition persist over time.

Finally, we deliberately chose a common-knowledge task for evaluation in our survey experiments because it aligned with our participants' likely knowledge and experience in our survey experiments, thus best mimicking a field experimental design. Although our field study among investment professionals certainly strengthens generalizability related to evaluator expertise and types of evaluation, future research can further expand on this work by studying misrecognition in different contexts.

In terms of the broader generalizability and scope conditions of our theory, it is important to note three conditions that may affect the magnitude of our observed effects: clarity of performance information, specificity of evaluative criteria, and visibility of evaluative outcomes

(for the self and others). Each of these conditions affect the awareness a person has that they have been misrecognized. For instance, the more concretely a person knows whether they were a top performer, that recognition ought to go to top performers, and that similarly performing others were recognized, the more likely it is that a top performer who is not recognized experiences underrecognition. Similarly, if a person who suspects they have been overrecognized can observe that a stronger performer was not recognized, we would expect them to be particularly aware of their overrecognition. Therefore, when these conditions are present—particularly when multiple of these hold—we would expect it to be *especially likely* that misrecognized evaluators subsequently misrecognize others.

It is worth noting that these conditions operate on a continuum—such that they are not simply “present or not present” but instead often present at varying degrees. Thus, while these conditions are sufficient, we argue they are *not necessary* for misrecognition to be reproduced in evaluations. For instance, an individual does not simply know or not know how they performed; rather the degree of clarity about how one performs varies across contexts and evaluative processes. Evaluators across all three of our studies had a reasonable sense of whether they were misrecognized. This was especially true in Study 1, given that participants in Study 2 and Study 3 were simply told whether they performed above or below average. Furthermore, Additional Study A was designed to make this performance distinction *even more ambiguous*. Relatedly, Additional Study B made the evaluative criteria more ambiguous. Results from these additional studies are consistent with our main results, and thus support the conjecture that misrecognized evaluators are apt to misrecognize others even in the absence of these conditions. A fruitful direction for future research is to more deeply consider how these (and other related) conditions

that commonly vary across evaluative contexts distinctly shape the reproduction of misrecognition.

Future research can also consider whether the effects of misrecognition carry across evaluation systems. We focused on peer evaluation systems because, in such processes, individuals who are evaluated typically also serve as evaluators. Furthermore, these types of evaluations are highly consequential and common in organizations, in science, and in society. Former candidates often also advance to evaluator positions in a different system, such as a manager being hired into a new organization (Castilla and Ranganathan 2020). Given that we find underrecognition and overrecognition to have clear and substantively large causal effects on how people evaluate others within the same system, it is plausible that the effects of misrecognition in one system can also shape evaluative behavior in another system. On the one hand, it is possible that overrecognized evaluators carry the insights they glean on how to evaluate into other systems, such that overrecognition has more enduring effects on how people later evaluate others. On the other hand, negative experiences, such as underrecognition, are commonly more salient (e.g., Baumeister et al. 2001), suggesting that these experiences may have more enduring effects. An opportunity thus exists for future research to examine the conditions under which misrecognition can spill over from one evaluative system (or context) to another and ascertain whether distinct mechanisms emerge in these cases. More generally, our findings caution against making strong claims that an evaluative process is meritocratic in cases where there is a reasonable risk of misrecognition. Such claims are apt to heighten expectations that outcomes will be meritocratic and make instances of misrecognition more salient.

In sum, this paper draws attention to how structural failures in evaluative processes can become reinforced as evaluators reproduce their experiences when evaluating others. Because

evaluators rely on their prior experiences of having been evaluated, inefficiencies in evaluative systems can have the tendency to be reproduced over time. The stakes of these misrecognitions are thus significant for organizations and our society at large. As we have documented, the differences in how misrecognized evaluators recognize peers occur in two critical ways: They dole out recognition at different rates *and* deviate from the general norm of recognizing only the strongest performers. And because underrecognized and overrecognized evaluators are sensitive to different mechanisms, it is imperative that in the design of evaluative systems, we better identify where breakdowns in meritocracy arise, how these breakdowns manifest differently for misrecognized individuals, and what the best solutions are to amend them.

REFERENCES

- Abelson, Robert P. 1981. "Psychological Status of the Script Concept." *American Psychologist* 36(7):715–29. doi: 10.1037/0003-066X.36.7.715.
- Abraham, Mabel. 2017. "Pay Formalization Revisited: Considering the Effects of Manager Gender and Discretion on Closing the Gender Wage Gap." *Academy of Management Journal* 60(1):29–54. doi: 10.5465/amj.2013.1060.
- Abraham, Mabel, Tristan L. Botelho, and Gabrielle Lamont-Dobbin. 2024. "The (Re)Production of Inequality in Evaluations: A Unifying Framework Outlining the Drivers of Gender and Racial Differences in Evaluative Outcomes." *Research in Organizational Behavior*.
- Accominotti, Fabien, Freda Lynn, and Michael Sauder. 2022. "The Architecture of Status Hierarchies: Variations in Structure and Why They Matter for Inequality." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 8(6):87–102. doi: 10.7758/RSF.2022.8.6.05.
- Adams, J. Stacy. 1965. "Inequity In Social Exchange." Pp. 267–99 in *Advances in Experimental Social Psychology*. Vol. 2, edited by L. Berkowitz. Academic Press.
- Adams, J. Stacy, and William B. Rosenbaum. 1962. "The Relationship of Worker Productivity to Cognitive Dissonance about Wage Inequities." *Journal of Applied Psychology* 46(3):161–64. doi: 10.1037/h0047751.
- Alon, Sigal, and Marta Tienda. 2007. "Diversity, Opportunity, and the Shifting Meritocracy in Higher Education." *American Sociological Review* 72(4):487–511.
- Ambrose, Maureen L., Mark A. Seabright, and Marshall Schminke. 2002. "Sabotage in the Workplace: The Role of Organizational Injustice." *Organizational Behavior and Human Decision Processes* 89(1):947–65. doi: 10.1016/S0749-5978(02)00037-7.
- Anderson, Cynthia D., and Donald Tomaskovic-Devey. 1995. "Patriarchal Pressures: An Exploration of Organizational Processes That Exacerbate and Erode Gender Earnings Inequality." *Work and Occupations* 22(3):328–56. doi: 10.1177/0730888495022003005.
- Austin, William, and Elaine Walster. 1974. "Reactions to Confirmations and Disconfirmations of Expectancies of Equity and Inequity." *Journal of Personality and Social Psychology* 30(2):208–16. doi: 10.1037/h0036622.
- Azoulay, Pierre, Toby Stuart, and Yanbo Wang. 2013. "Matthew: Effect or Fable?" *Management Science* 60(1):92–109. doi: 10.1287/mnsc.2013.1755.
- Baumeister, Roy F., Ellen Bratslavsky, Catrin Finkenauer, and Kathleen D. Vohs. 2001. "Bad Is Stronger than Good." *Review of General Psychology* 5(4):323–70. doi: 10.1037/1089-2680.5.4.323.

- Baumeister, Roy F., Arlene M. Stillwell, and Todd F. Heatherton. 1994. "Guilt: An Interpersonal Approach." *Psychological Bulletin* 115(2):243–67. doi: 10.1037/0033-2909.115.2.243.
- Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review* 95(3):616–36. doi: 10.1257/0002828054201413.
- Bian, Jiang, Jason Greenberg, Jizhen Li, and Yanbo Wang. 2022. "Good to Go First? Position Effects in Expert Evaluation of Early-Stage Ventures." *Management Science* 68(1):300–315. doi: 10.1287/mnsc.2021.4132.
- Biddle, B. J. 1986. "Recent Developments in Role Theory." *Annual Review of Sociology* 12(1):67–92. doi: 10.1146/annurev.so.12.080186.000435.
- Blau, Peter Michael. 1964. *Exchange and Power in Social Life*. J. Wiley.
- Bol, Thijs, Mathijs de Vaan, and Arnout van de Rijt. 2018. "The Matthew Effect in Science Funding." *Proceedings of the National Academy of Sciences* 115(19):4887–90. doi: 10.1073/pnas.1719557115.
- Bond, Brittany M. 2020. "Essays on Status Recognition and Its Consequences for Top-Talent Mobility and Productivity." Thesis, Massachusetts Institute of Technology.
- Bond, Brittany M. 2024. "Cut to the Curve: Underrecognition and Talent Loss for Forced Ranking in a Multinational Firm." *Working Paper*.
- van den Bos, Kees, Henk A. M. Wilke, E. Allan Lind, and Riël Vermunt. 1998. "Evaluating Outcomes by Means of the Fair Process Effect: Evidence for Different Processes in Fairness and Satisfaction Judgments." *Journal of Personality and Social Psychology* 74(6):1493–1503. doi: 10.1037/0022-3514.74.6.1493.
- Botelho, Tristan L. 2024. "From Audience to Evaluator: When Visibility into Prior Evaluations Leads to Convergence or Divergence in Subsequent Evaluations Among Professionals." *Organization Science*. doi: 10.1287/orsc.2017.11285.
- Botelho, Tristan L., and Mabel Abraham. 2017. "Pursuing Quality: How Search Costs and Uncertainty Magnify Gender-Based Double Standards in a Multistage Evaluation Process." *Administrative Science Quarterly* 62(4):698–730. doi: 10.1177/0001839217694358.
- Botelho, Tristan L., and Marina Gertsberg. 2021. "The Disciplining Effect of Status: Evaluator Status Awards and Observed Gender Bias in Evaluations." *Management Science*. doi: 10.1287/mnsc.2021.4161.
- Bourdieu, Pierre. 1993. *The Field of Cultural Production*. Columbia University Press.
- Bowers, Anne, and Matteo Prato. 2018. "The Structural Origins of Unearned Status: How Arbitrary Changes in Categories Affect Status Position and Market Impact." *Administrative Science Quarterly* 63(3):668–99. doi: 10.1177/0001839217727706.

- Brett, Joan F., and Leanne E. Atwater. 2001. "360° Feedback: Accuracy, Accuracy, Reactions, and Perceptions of Usefulness." *Journal of Applied Psychology* 86(5):930–42. doi: 10.1037/0021-9010.86.5.930.
- Brockner, Joel. 1985. "The Relation of Trait Self-Esteem and Positive Inequity to Productivity." *Journal of Personality* 53(4):517–29. doi: 10.1111/j.1467-6494.1985.tb00380.x.
- Brockner, Joel, Mary Konovsky, Rochelle Cooper-Schneider, Robert Folger, Christopher Martin, and Robert J. Bies. 1994. "Interactive Effects of Procedural Justice and Outcome Negativity on Victims and Survivors of Job Loss." *The Academy of Management Journal* 37(2):397–409. doi: 10.2307/256835.
- Campanario, Juan Miguel. 1998. "Peer Review for Journals as It Stands Today—Part 1." *Science Communication* 19(3):181–211. doi: 10.1177/1075547098019003002.
- Campbell, Donald T., and Julian Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. 1st edition. Belmont, CA: Cengage Learning.
- Carrell, Michael R., and John E. Dittrich. 1978. "Equity Theory: The Recent Literature, Methodological Considerations, and New Directions." *The Academy of Management Review* 3(2):202–10. doi: 10.2307/257661.
- Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 113(6):1479–1526. doi: 10.1086/517895.
- Castilla, Emilio J. 2015. "Accounting for the Gap: A Firm Study Manipulating Organizational Accountability and Transparency in Pay Decisions." *Organization Science* 26(2):311–33. doi: 10.1287/orsc.2014.0950.
- Castilla, Emilio J., and Stephen Benard. 2010. "The Paradox of Meritocracy in Organizations." *Administrative Science Quarterly* 55(4):543–76.
- Castilla, Emilio J., and Aruna Ranganathan. 2020. "The Production of Merit: How Managers Understand and Apply Merit in the Workplace." *Organization Science* 31(4):909–35. doi: 10.1287/orsc.2019.1335.
- Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer. 2018. *Loss Attitudes in the U.S. Population: Evidence from Dynamically Optimized Sequential Experimentation (DOSE)*. Working Paper. 25072. National Bureau of Economic Research. doi: 10.3386/w25072.
- Cialdini, Robert B., and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55:591–621. doi: <http://dx.doi.org/10.1146/annurev.psych.55.090902.142015>.
- Cohen, Lisa E., Joseph P. Broschak, and Heather A. Haveman. 1998. "And Then There Were More? The Effect of Organizational Sex Composition on the Hiring and Promotion of Managers." *American Sociological Review* 63(5):711–27. doi: 10.2307/2657335.

- Cohen, Philip N., and Matt L. Huffman. 2007. "Working for the Woman? Female Managers and the Gender Wage Gap." *American Sociological Review* 72(5):681–704. doi: 10.1177/000312240707200502.
- Cohen-Charash, Yochi, and Paul E. Spector. 2001. "The Role of Justice in Organizations: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 86(2):278–321. doi: 10.1006/obhd.2001.2958.
- Cook, Karen S., and Karen A. Hegtvedt. 1983. "Distributive Justice, Equity, and Equality." *Annual Review of Sociology* 9(1):217–41. doi: 10.1146/annurev.so.09.080183.001245.
- Correll, Shelley J., Stephen Benard, and In Paik. 2007. "Getting a Job: Is There a Motherhood Penalty?" *The American Journal of Sociology* 112(5):1297–1338.
- Correll, Shelley J., Katherine R. Weisshaar, Alison T. Wynn, and JoAnne Delfino Wehner. 2020. "Inside the Black Box of Organizational Life: The Gendered Language of Performance Assessment." *American Sociological Review* 85(6):1022–50. doi: 10.1177/0003122420962080.
- Cropanzano, Russell, and Maureen L. Ambrose. 2001. "Procedural and Distributive Justice Are More Similar than You Think: A Monistic Perspective and a Research Agenda." Pp. 119–51 in *Advances in organization justice*. Stanford University Press.
- Dencker, John C. 2009. "Relative Bargaining Power, Corporate Restructuring, and Managerial Incentives." *Administrative Science Quarterly* 54(3):453–85.
- Dobbin, Frank. 2009. *Inventing Equal Opportunity*. Princeton, NJ: Princeton University Press.
- Dobbin, Frank, Daniel Schrage, and Alexandra Kalev. 2015. "Rage against the Iron Cage: The Varied Effects of Bureaucratic Personnel Reforms on Diversity." *American Sociological Review* 80(5):1014–44.
- Dworak, Elizabeth M., William Revelle, Philip Doebler, and David M. Condon. 2021. "Using the International Cognitive Ability Resource as an Open Source Tool to Explore Individual Differences in Cognitive Ability." *Personality and Individual Differences* 169:109906. doi: 10.1016/j.paid.2020.109906.
- Elvira, Marta M., and Mary E. Graham. 2002. "Not Just a Formality: Pay System Formalization and Sex-Related Earnings Effects." *Organization Science* 13(6):601–17.
- Espenshade, Thomas J., and Alexandria Walton Radford. 2009. "No Longer Separate, Not Yet Equal: Race and Class in Elite College Admission and Campus Life." in *No Longer Separate, Not Yet Equal*. Princeton University Press.
- Fini, Riccardo, Julien Jourdan, Markus Perkmann, and Laura Toschi. 2022. "A New Take on the Categorical Imperative: Gatekeeping, Boundary Maintenance, and Evaluation Penalties in Science." *Organization Science*. doi: 10.1287/orsc.2022.1610.

- Foschi, Martha. 2000. "Double Standards for Competence: Theory and Research." *Annual Review of Sociology* 26(1):21–42. doi: 10.1146/annurev.soc.26.1.21.
- Gallus, Jana, and Bruno S. Frey. 2016. "Awards: A Strategic Management Perspective." *Strategic Management Journal* 37(8):1699–1714. doi: 10.1002/smj.2415.
- Gaucher, Danielle, Carolyn L. Hafer, Aaron C. Kay, and Nicolas Davidenko. 2010. "Compensatory Rationalizations and the Resolution of Everyday Undeserved Outcomes." *Personality and Social Psychology Bulletin* 36(1):109–18. doi: 10.1177/0146167209351701.
- Golden-Biddle, Karen, and Hayagreeva Rao. 1997. "Breaches in the Boardroom: Organizational Identity and Conflicts of Commitment in a Nonprofit Organization." *Organization Science* 8(6):593–611.
- Goode, William J. 1960. "A Theory of Role Strain." *American Sociological Review* 25(4):483–96.
- Greenberg, Jason, and Ethan Mollick. 2017. "Activist Choice Homophily and the Crowdfunding of Female Founders." *Administrative Science Quarterly* 62(2):341–74. doi: 10.1177/0001839216678847.
- Greenberg, Jerald. 1988. "Equity and Workplace Status: A Field Experiment." *Journal of Applied Psychology* 73(4):606–13. doi: 10.1037/0021-9010.73.4.606.
- Hahl, Oliver, Minjae Kim, and Ezra W. Zuckerman Sivan. 2018. "The Authentic Appeal of the Lying Demagogue: Proclaiming the Deeper Truth about Political Illegitimacy." *American Sociological Review* 83(1):1–33. doi: 10.1177/0003122417749632.
- Harder, Joseph W. 1992. "Play for Pay: Effects of Inequity in a Pay-for-Performance Context." *Administrative Science Quarterly* 37(2):321–35. doi: 10.2307/2393227.
- Homans, George C. 1951. *The Human Group*. Piscataway, NJ, US: Transaction Publishers.
- Huang, Laura, and Jone L. Pearce. 2015. "Managing the Unknowable: The Effectiveness of Early-Stage Investor Gut Feel in Entrepreneurial Investment Decisions." *Administrative Science Quarterly* 60(4):634–70. doi: 10.1177/0001839215597270.
- Hurwitz, Michael. 2011. "The Impact of Legacy Status on Undergraduate Admissions at Elite Colleges and Universities." *Economics of Education Review* 30(3):480–92. doi: 10.1016/j.econedurev.2010.12.002.
- Jensen, Michael, Torsten Twardawski, and Nadja Younes. 2021. "The Paradox of Awards: How Status Ripples Affect Who Benefits from CEO Awards." *Organization Science*. doi: 10.1287/orsc.2021.1475.

- Jost, John T., and Jojanneke van der Toorn. 2012. "System Justification Theory." Pp. 313–43 in *Handbook of theories of social psychology, Vol. 2*. Thousand Oaks, CA: Sage Publications Ltd.
- Kenny, David. 2019. "Enhancing Validity in Psychological Research." *The American Psychologist* 74:1018–28. doi: 10.1037/amp0000531.
- Kidd, Robert F., and Mary K. Utne. 1978. "Reactions to Inequity: A Prospective on the Role of Attributions." *Law and Human Behavior* 2(4):301–12. doi: 10.1007/BF01038983.
- Kim, Minjae, and Daniel DellaPosta. 2022. "The Fickle Crowd: Reinforcement and Contradiction of Quality Evaluations in Cultural Markets." *Organization Science* 33(6):2496–2518. doi: 10.1287/orsc.2021.1556.
- Kim Jerry W., King Brayden G. 2014. "Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring." *Management Science* 60(11):2619–44.
- Klapper, Helge, Henning Piezunka, and Linus Dahlander. 2024. "Peer Evaluations: Evaluating and Being Evaluated." *Organization Science* 35(4):1363–87. doi: 10.1287/orsc.2021.15302.
- Konovsky, Mary A. 2000. "Understanding Procedural Justice and Its Impact on Business Organizations." *Journal of Management* 26(3):489–511. doi: 10.1016/S0149-2063(00)00042-8.
- Kuwabara, Ko. 2015. "Do Reputation Systems Undermine Trust? Divergent Effects of Enforcement Type on Generalized Trust and Trustworthiness." *American Journal of Sociology* 120(5):1390–1428. doi: 10.1086/681231.
- Lamont, Michèle. 2012. "Toward a Comparative Sociology of Valuation and Evaluation." *Annual Review of Sociology* 38(1):201–21. doi: 10.1146/annurev-soc-070308-120022.
- Lawler, Edward E. 1968. "Effects of Hourly Overpayment on Productivity and Work Quality." *Journal of Personality and Social Psychology* 10(3):306–14. doi: 10.1037/h0026614.
- Lerner, Melvin J. 1980. "The Belief in a Just World." Pp. 9–30 in *The Belief in a Just World: A Fundamental Delusion, Perspectives in Social Psychology*, edited by M. J. Lerner. Boston, MA: Springer US.
- Leung, Ming D., and Sharon Koppman. 2018. "Taking a Pass: How Proportional Prejudice and Decisions Not to Hire Reproduce Gender Segregation." *American Journal of Sociology* 124(3):762–813. doi: 10.1086/700677.
- Liberman, Nira, Lorraine Chen Idson, and E. Tory Higgins. 2005. "Predicting the Intensity of Losses vs. Non-Gains and Non-Losses vs. Gains in Judging Fairness and Value: A Test of the Loss Aversion Explanation." *Journal of Experimental Social Psychology* 41(5):527–34. doi: 10.1016/j.jesp.2004.06.007.

- Lind, E. Allan. 2001. "Fairness Heuristic Theory: Justice Judgments as Pivotal Cognitions in Organizational Relations." Pp. 56–88 in *Advances in organization justice*. Stanford University Press.
- Lind, E. Allan, and Tom R. Tyler. 1988. *The Social Psychology of Procedural Justice*. Boston, MA: Springer US.
- Litman, Leib, Jonathan Robinson, and Tzvi Abberbock. 2017. "TurkPrime.Com: A Versatile Crowdsourcing Data Acquisition Platform for the Behavioral Sciences." *Behavior Research Methods* 49(2):433–42. doi: 10.3758/s13428-016-0727-z.
- Lucas, Brian J., Zachariah Berry, Laura M. Giurge, and Dolly Chugh. 2021. "A Longer Shortlist Increases the Consideration of Female Candidates in Male-Dominant Domains." *Nature Human Behaviour* 5(6):736–42. doi: 10.1038/s41562-020-01033-0.
- Lynn, Freda, Joel Podolny, and Lin Tao. 2009. "A Sociological (De)Construction of the Relationship between Status and Quality." *American Journal of Sociology* 115:755–804. doi: 10.1086/603537.
- McNamee, Stephen J., and Robert K. Miller. 2009. *The Meritocracy Myth, 2nd Ed.* Lanham, MD, US: Rowman & Littlefield.
- Merton, Robert K. 1968. "The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered." *Science* 159(3810):56–63. doi: 10.1126/science.159.3810.56.
- Mijs, Jonathan J. B. 2018. "Visualizing Belief in Meritocracy, 1930–2010." *Socius* 4:2378023118811805. doi: 10.1177/2378023118811805.
- Miles, Andrew, Raphaël Charron-Chénier, and Cyrus Schleifer. 2019. "Measuring Automatic Cognition: Advancing Dual-Process Research in Sociology." *American Sociological Review* 84(2):308–33. doi: 10.1177/0003122419832497.
- Mobasser, Sanaz. 2019. "Race, Place, and Crime: How Violent Crime Events Affect Employment Discrimination." *American Journal of Sociology* 125(1):63–104. doi: 10.1086/703883.
- Ody-Brasier, Amandine, and Amanda Sharkey. 2019. "Under Pressure: Reputation, Ratings, and Inaccurate Self-Reporting in the Nursing Home Industry." *Strategic Management Journal* 40(10):1517–44. doi: 10.1002/smj.3063.
- Paik, Eugene Taeha, Timothy G. Pollock, Steven Boivie, Donald Lange, and Peggy M. Lee. 2023. "A Star Is Born: The Relationship Between Performance and Achieving Status Through Certification Contests in the Context of Equity Analysts." *Organization Science* 34(1):75–99. doi: 10.1287/orsc.2021.1563.

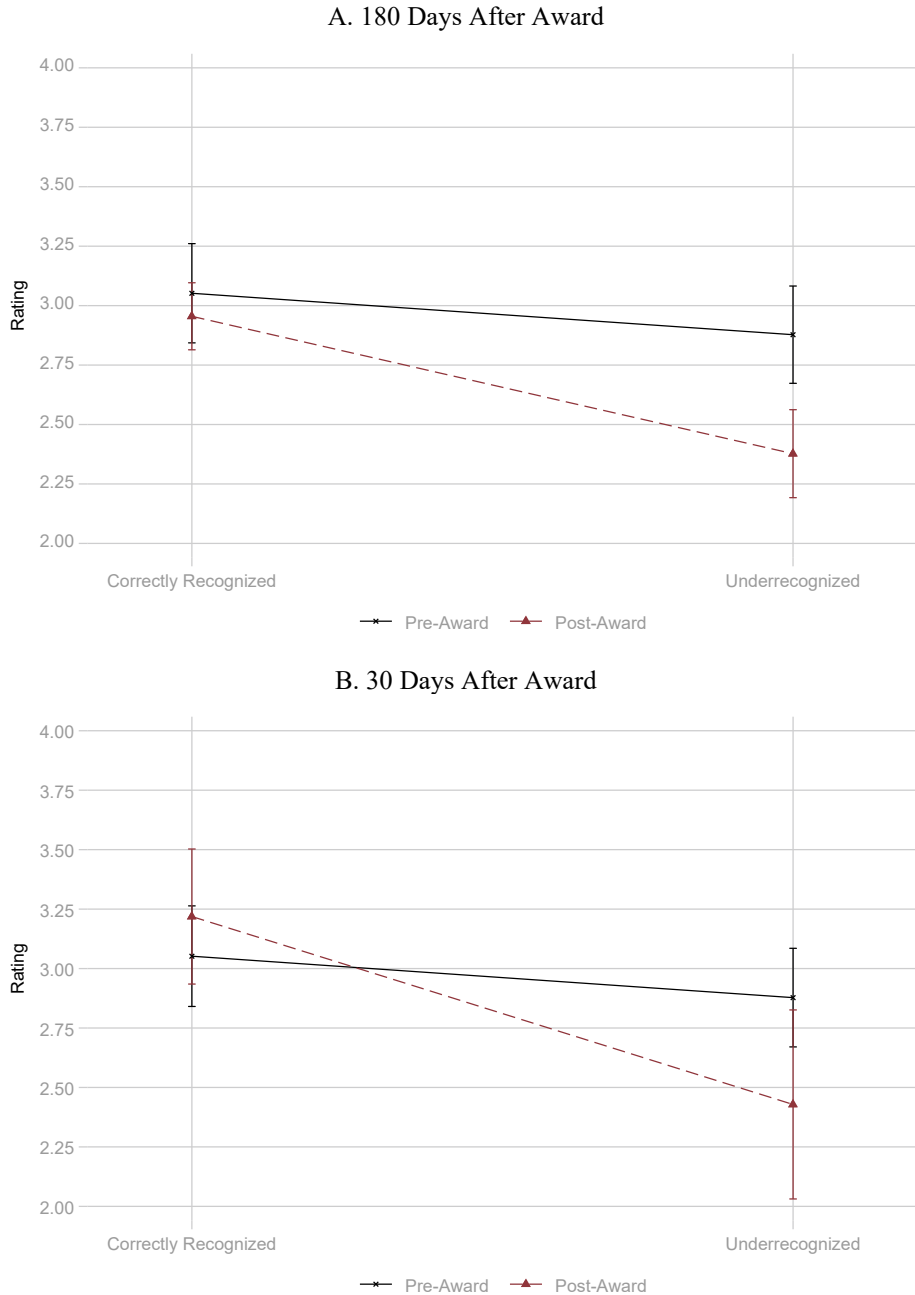
- Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5(5):411–19. doi: 10.1017/S1930297500002205.
- Paxton, Pamela, Kristopher Velasco, and Robert W. Ressler. 2020. "Does Use of Emotion Increase Donations and Volunteers for Nonprofits?" *American Sociological Review* 85(6):1051–83. doi: 10.1177/0003122420960104.
- Pedulla, David S. 2016. "Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories." *American Sociological Review* 81(2):262–89. doi: 10.1177/0003122416630982.
- Podolny, Joel M. 1993. "A Status-Based Model of Market Competition." *The American Journal of Sociology* 98(4):829–72.
- Pritchard, Robert D., Marvin D. Dunnette, and Dale O. Gorgenson. 1972. "Effects of Perceptions of Equity and Inequity on Worker Performance and Satisfaction." *Journal of Applied Psychology* 56(1):75–94. doi: 10.1037/h0032135.
- Rahman, Hatim A. 2021. "The Invisible Cage: Workers' Reactivity to Opaque Algorithmic Evaluations." *Administrative Science Quarterly* 66(4):945–88. doi: 10.1177/00018392211010118.
- Reskin, Barbara F. 2000. "The Proximate Causes of Employment Discrimination." *Contemporary Sociology* 29(2):319–28. doi: 10.2307/2654387.
- Ridgeway, Cecilia L., and Shelley J. Correll. 2006. "Consensus and the Creation of Status Beliefs." *Social Forces* 85(1):431–53. doi: 10.2307/3844422.
- Rivera, Alba N., and James T. Tedeschi. 1976. "Public versus Private Reactions to Positive Inequity." *Journal of Personality and Social Psychology* 34(5):895–900. doi: 10.1037/0022-3514.34.5.895.
- Rivera, Lauren A. 2015. "Go with Your Gut: Emotion and Evaluation in Job Interviews." *American Journal of Sociology* 120(5):1339–89. doi: 10.1086/681214.
- Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84(2):248–74. doi: 10.1177/0003122419833601.
- Saguy, Tamar, Saulo Fernández, Nyla R. Branscombe, and Aviv Shany. 2020. "Justice Agents: Discriminated Group Members Are Perceived to Be Highly Committed to Social Justice." *Personality and Social Psychology Bulletin* 46(1):155–67. doi: 10.1177/0146167219845922.
- Scheer, Lisa K., Nirmalya Kumar, and Jan-Benedict E. M. Steenkamp. 2003. "Reactions to Perceived Inequity in U.S. and Dutch Interorganizational Relationships." *The Academy of Management Journal* 46(3):303–16. doi: 10.2307/30040624.

- Scully, M. A. 2000. "Manage Your Own Employability: Meritocracy and the Legitimation of Inequality in Internal Labor Markets and Beyond." Pp. 199–214 in *Relational wealth: The advantages of stability in a changing economy*. Oxford, UK: Oxford University Press.
- Sharkey, Amanda J., and Balázs Kovács. 2017. "The Many Gifts of Status: How Attending to Audience Reactions Drives the Use of Status." *Management Science*. doi: 10.1287/mnsc.2017.2879.
- Shaw, Alex, and Kristina R. Olson. 2012. "Children Discard a Resource to Avoid Inequity." *Journal of Experimental Psychology: General* 141(2):382–95. doi: 10.1037/a0025907.
- Siler, Kyle, Kirby Lee, and Lisa Bero. 2015. "Measuring the Effectiveness of Scientific Gatekeeping." *Proceedings of the National Academy of Sciences* 112(2):360–65. doi: 10.1073/pnas.1418218112.
- Simcoe, Timothy S., and Dave M. Waguespack. 2010. "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science* 57(2):274–90. doi: 10.1287/mnsc.1100.1270.
- Smith, Sandra Susan. 2005. "'Don't Put My Name on It': Social Capital Activation and Job-Finding Assistance among the Black Urban Poor." *American Journal of Sociology* 111(1):1–57. doi: 10.1086/428814.
- Sutton, John R., Frank Dobbin, John W. Meyer, and W. Richard Scott. 1994. "The Legalization of the Workplace." *American Journal of Sociology* 99(4):944–71. doi: 10.1086/230368.
- Sweeney, Paul D. 1990. "Distributive Justice and Pay Satisfaction: A Field Test of an Equity Theory Prediction." *Journal of Business and Psychology* 4(3):329–41. doi: 10.1007/BF01125243.
- Thye, Shane R. 2000. "Reliability in Experimental Sociology*." *Social Forces* 78(4):1277–1309. doi: 10.1093/sf/78.4.1277.
- Tilcsik, András. 2021. "Statistical Discrimination and the Rationalization of Stereotypes." *American Sociological Review* 86(1):93–122. doi: 10.1177/0003122420969399.
- Uhlmann, Eric Luis, and Geoffrey L. Cohen. 2005. "Constructed Criteria: Redefining Merit to Justify Discrimination." *Psychological Science* 16(6):474–80. doi: 10.1111/j.0956-7976.2005.01559.x.
- Veen, Alex, Tom Barratt, and Caleb Goods. 2020. "Platform-Capital's 'App-Etite' for Control: A Labour Process Analysis of Food-Delivery Work in Australia." *Work, Employment and Society* 34(3):388–406. doi: 10.1177/0950017019836911.
- Vining, Daniel R. 1984. "Subfertility among the Very Intelligent: An Examination of the American Mensa." *Personality and Individual Differences* 5(6):725–33. doi: 10.1016/0191-8869(84)90120-X.

- Walster, Elaine, and Jane Allyn Piliavin. 1972. "Equity and the Innocent Bystander." *Journal of Social Issues* 28(3):165–89. doi: 10.1111/j.1540-4560.1972.tb00038.x.
- Weiss, Howard M., Kathleen Suckow, and Russell Cropanzano. 1999. "Effects of Justice Conditions on Discrete Emotions." *Journal of Applied Psychology* 84(5):786–94. doi: 10.1037/0021-9010.84.5.786.
- Weisshaar, Katherine. 2017. "Publish and Perish? An Assessment of Gender Gaps in Promotion to Tenure in Academia." *Social Forces* 96(2):529–60. doi: 10.1093/sf/sox052.
- Yang, Liu-Qin, Jeremy Bauer, Russell E. Johnson, Maureen W. Groer, and Kristen Salomon. 2014. "Physiological Mechanisms That Underlie the Effects of Interactional Unfairness on Deviant Behavior: The Role of Cortisol Activity." *The Journal of Applied Psychology* 99(2):310–21. doi: 10.1037/a0034413.
- Yang, Tiantian, and Howard E. Aldrich. 2014. "Who's the Boss? Explaining Gender Inequality in Entrepreneurial Teams." *American Sociological Review* 79(2):303–27. doi: 10.1177/0003122414524207.
- Zhou, Xiaolin, and Yan Wu. 2011. "Sharing Losses and Sharing Gains: Increased Demand for Fairness under Adversity." *Journal of Experimental Social Psychology* 47(3):582–88. doi: 10.1016/j.jesp.2010.12.017.

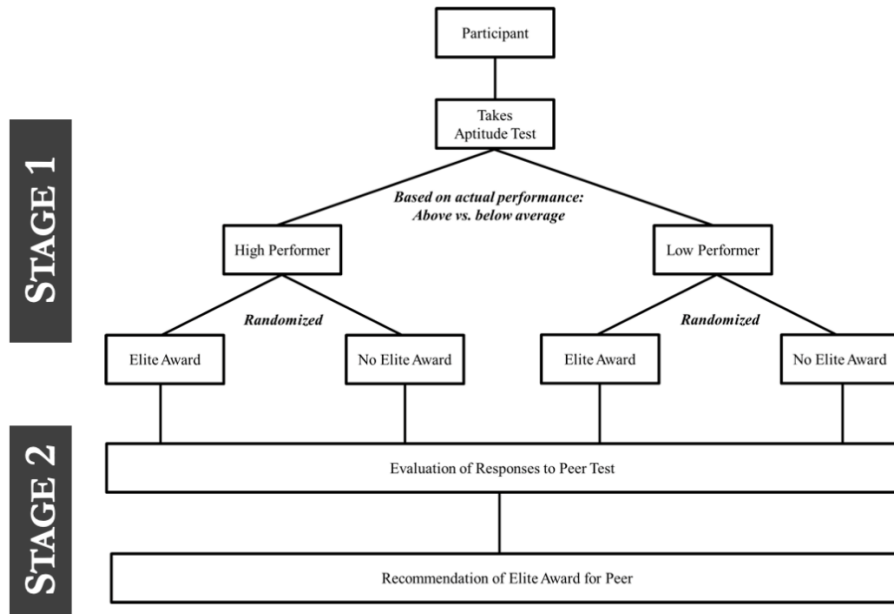
FIGURES AND TABLES

Figure 1. Comparing Average Rating Between Correctly Recognized and Underrecognized (Study 1)



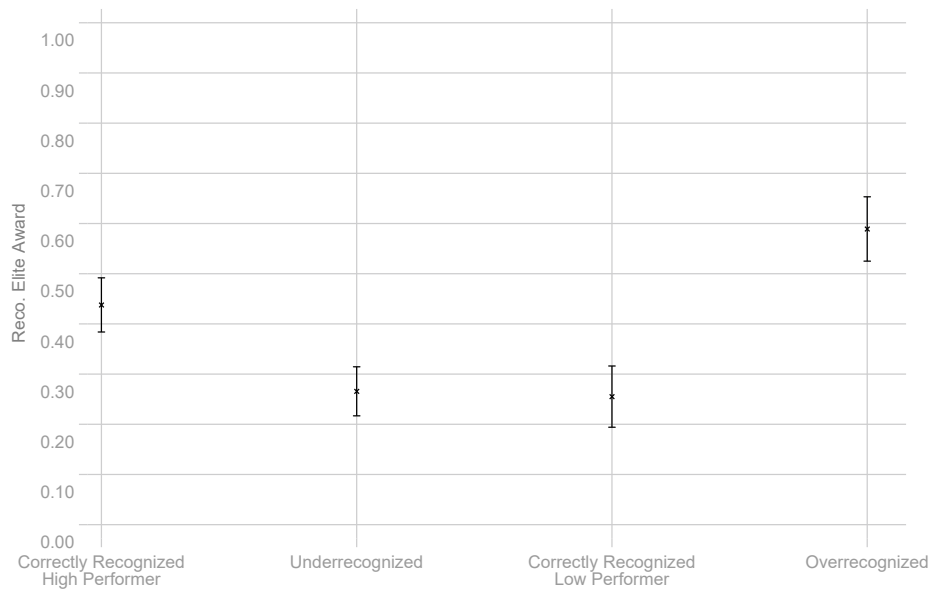
Note: This figure depicts the average value of *Rating* (1-5). Each week a professional was recognized with a platform wide e-mail recognizing the professional and their recent investment recommendation. The sample includes all weeks where two professionals were equally eligible but were chosen in a quasi-random fashion. *Correctly Recognized* represents the chosen investment professional and *Underrecognized* represents the equally qualified professional who was not chosen. *Pre-Award* represents the average rating value given by professionals before the award e-mail and *Post-Award* represents the average rating value given by professionals after the award e-mail. Figure 1A limits the sample to 180 days before and after the award and Figure 1B limits the sample to the 180 days before and 30 days after the award.

Figure 2. Experimental Design (Study 2)



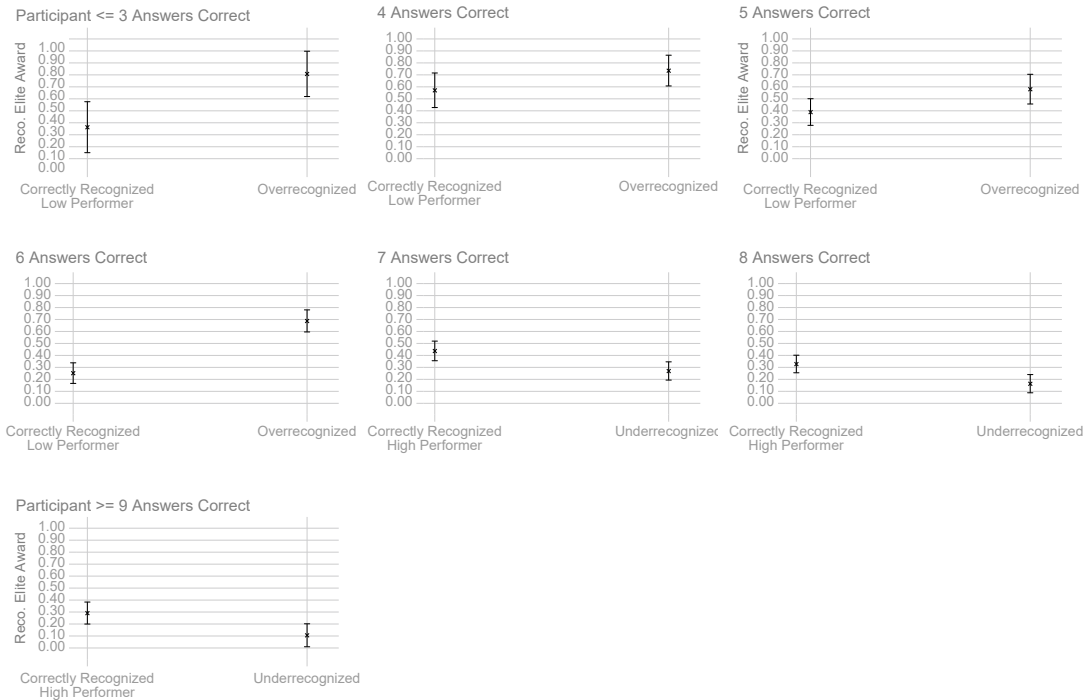
Note: This figure depicts the experimental design of Study 2.

Figure 3. Participants' Propensity to Recommend Elite Award, by Condition, with Controls (Study 2)



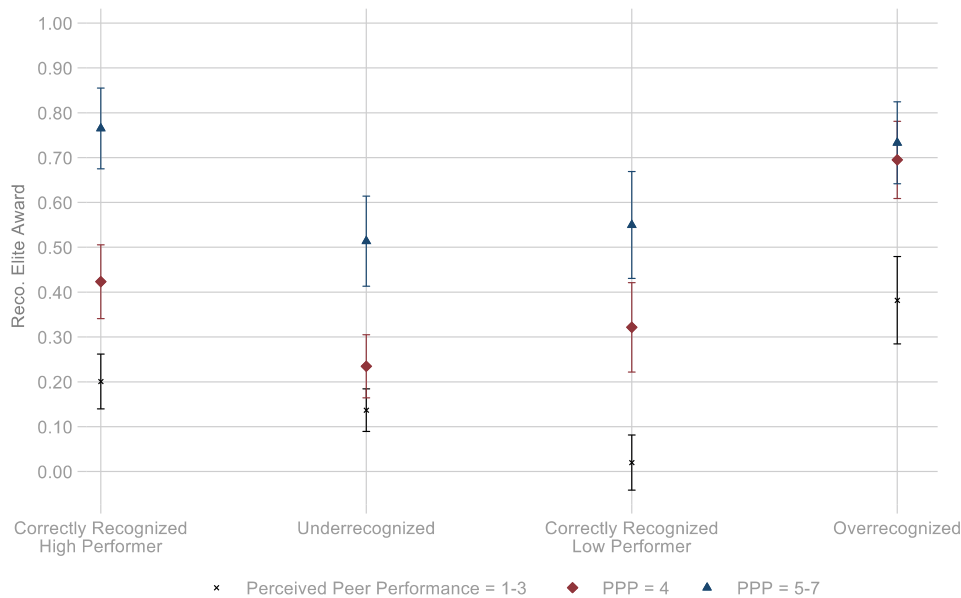
Note: This figure shows marginal effects from a regression of *Recommend Elite Award*, by condition, with controls (Table 2, Model 1). Bars represent 95 percent confidence intervals.

Figure 4. Participants' Propensity to Recommend Elite Award, by Number of Questions they Answered Correctly (Study 2)



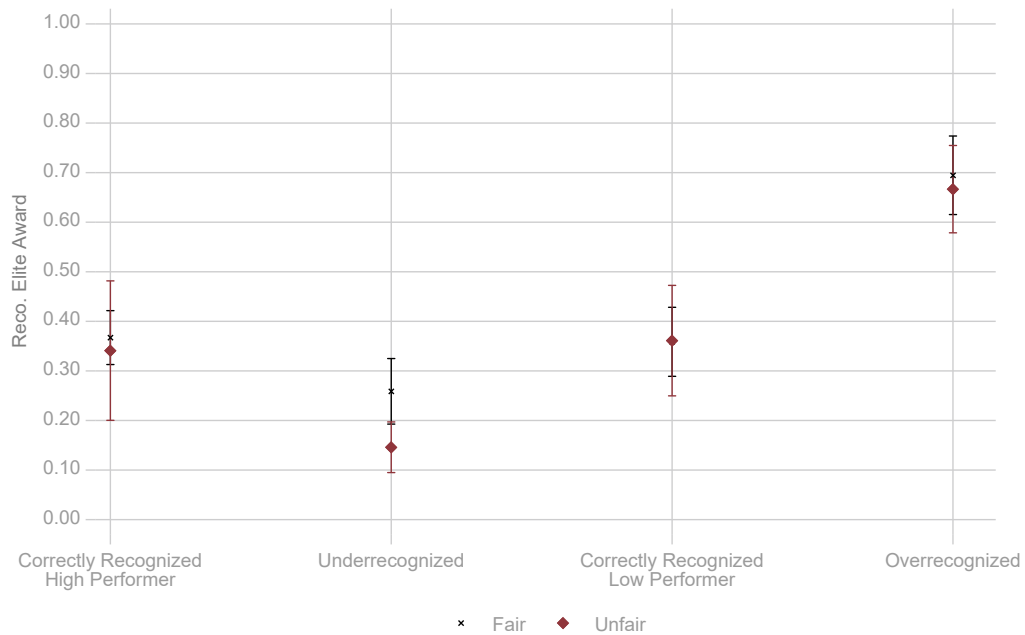
Note: This figure shows the mean of *Recommend Elite Award*, by the number of assessment questions the participant answered correctly and by condition. Participants who answered two or fewer questions are grouped with those answering three questions correctly and participants who answered ten questions correctly are grouped with those answering nine questions correctly, given the rarity (1.6 percent and 3.1 percent, respectively). Bars represent 95 percent confidence intervals.

Figure 5. Participants' Perception of Perceived Peer Performance and Recommending Elite Award (Study 2)



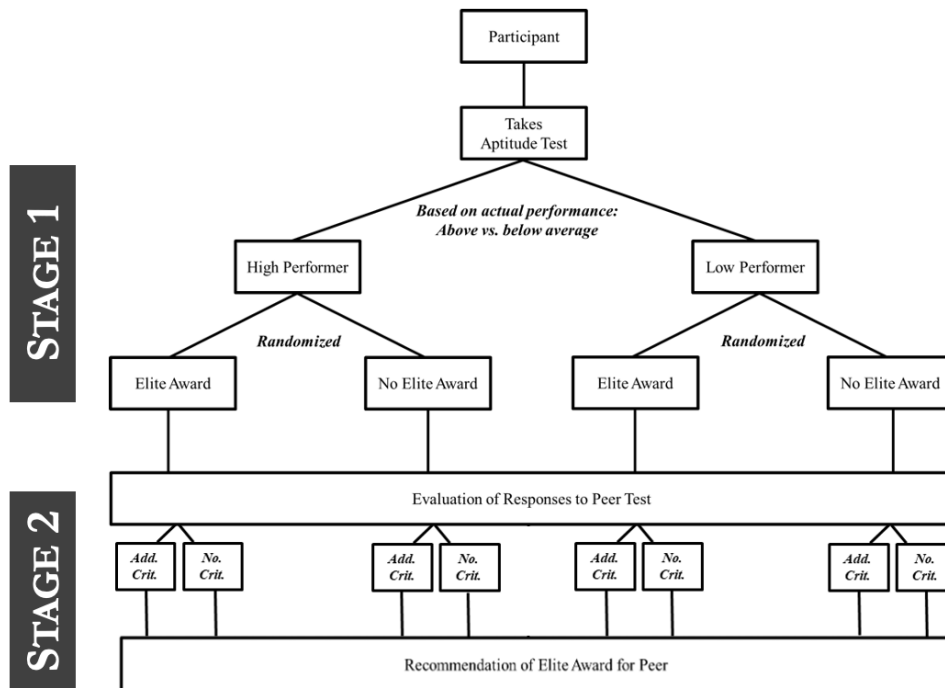
Note: This figure shows marginal effects from a regression of *Recommend Elite Award* on an interaction between condition and *Perceived Peer Performance* with controls (Table C5).

Figure 6. Participants' Perception of Fairness and Recommending Elite Award (Study 2)



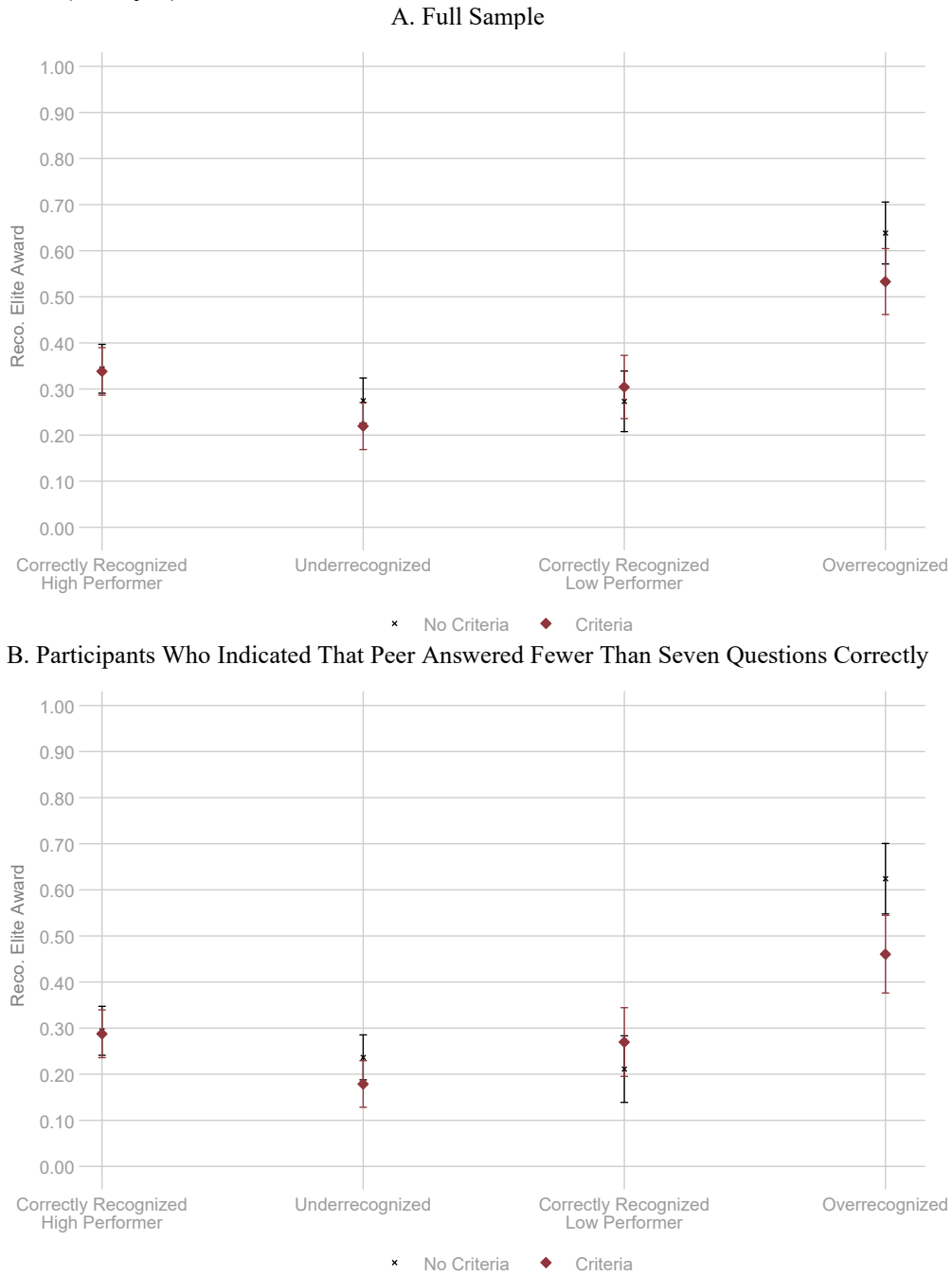
Note: This figure shows marginal effects from a regression of *Recommend Elite Award* on an interaction between condition and *Unfair* (Table 2, Model 2). *Unfair* takes the value of 1 if the participant responded that the allocation of the Elite Award was unfair (values 1-3 out of 7). Bars represent 95 percent confidence intervals.

Figure 7. Experimental Design for Study 3



Note: This figure depicts the experimental design of Study 3. In addition to being randomized to receive the Elite Award or not, participants were randomized into the “additional criteria” and “no criteria” branches.

Figure 8. Participants' Propensity to Recommend Elite Award by Condition and Criteria, with Controls (Study 3)



Note: This figure shows marginal effects from a regression of *Recommend Elite Award* on an interaction between condition and *Criteria* (Table 3). *Criteria* takes the value of 1 if the participant was randomly assigned to receive more information about allocating the Elite Award and 0 otherwise. Figure 8A is the full sample of participants (Table 3, Model 1) and Figure 8B limits the sample to participants who indicated that the peer answered fewer than seven questions correctly (Table 3, Model 2).

Table 1. Descriptive Statistics (Study 2)

Variables	N	Mean	SD	Min	Max
Recommend Elite Award	1,202	0.379	0.485	0.000	1.000
Perceived Peer Performance	1,202	3.891	0.999	1.000	7.000
Fairness Receiving Award	1,202	4.136	1.668	1.000	7.000
Positive Affect	1,202	4.162	1.310	1.000	7.000
Negative Affect	1,202	2.072	1.070	1.000	6.900
Participant Test Number Correct	1,202	6.669	1.788	0.000	10.000
Participant Test Time Spent	1,202	502.288	308.985	45.927	3295.744
Time Spent on Peer Test	1,202	248.795	173.741	14.832	1536.689
Accuracy on Peer Test	1,202	0.761	0.155	0.100	1.000
Number Marked Correct on Peer Test	1,202	4.837	1.518	0.000	10.000
Age	1,202	43.559	10.769	29.000	82.000
College Educated	1,202	0.572	0.495	0.000	1.000
Experience Cognitive Assessments	1,202	0.146	0.354	0.000	1.000
Female	1,202	0.573	0.495	0.000	1.000
Democrat	1,202	0.468	0.499	0.000	1.000
Republican	1,202	0.196	0.397	0.000	1.000
White	1,202	0.744	0.437	0.000	1.000
Full-Time Employee	1,202	0.581	0.494	0.000	1.000

Note: We measured positive and negative affect as exploratory variables and report the means by condition in the Appendix C, Table C6.

Table 2. OLS Regressions of Recommend Elite by Evaluators' Recognition and Fairness (Study 2)

	Model 1	Model 2
Underrecognized	-0.172*** (0.033)	-0.108* (0.044)
Correctly Recognized Low Performer	-0.183*** (0.048)	-0.009 (0.045)
Overrecognized	0.151** (0.049)	0.327*** (0.049)
Unfair		-0.026 (0.077)
Underrecognized × Unfair		-0.087 (0.088)
Correctly Recognized Low Performer × Unfair		0.029 (0.102)
Overrecognized × Unfair		-0.002 (0.098)
Perceived Peer Performance	0.175*** (0.013)	
Time Spent on Peer Test	-0.000 (0.000)	
Number Marked Correct on Peer Test	0.074*** (0.009)	
Accuracy on Peer Test	-0.220* (0.086)	
Fairness Receiving Award	-0.002 (0.008)	
Positive Affect	-0.006 (0.010)	
Negative Affect	0.004 (0.012)	
Participant Test Number Correct	-0.037** (0.012)	
Participant Test Time Spent	0.000* (0.000)	
Age	-0.001 (0.001)	
College Educated	-0.063** (0.023)	
Experience Cognitive Assessments	0.042 (0.032)	
Female	-0.048* (0.023)	
Democrat	0.030 (0.026)	
Republican	0.024 (0.031)	
White	-0.010 (0.027)	
Full-Time Employee	-0.005 (0.024)	
Constant	-0.095 (0.144)	0.367*** (0.028)
Observations	1,202	1,202
R-Squared Adj.	0.384	0.118

Note: Unit of analysis is the participant's evaluation of peer test. *Recommend Elite Award* takes the value of 1 if the participant recommended that the peer receive the Elite Award and 0 if they did not recommend it. *Underrecognized* takes the value of 1 if the participant performed above average on their aptitude test and was not randomly assigned the Elite Award and 0 otherwise. *Correctly Recognized Low Performer* takes the value of 1 if the participant performed below average on their aptitude test and was not randomly assigned the Elite Award and 0 otherwise. *Overrecognized* takes the value of 1 if the participant performed below average on their aptitude test and was randomly assigned the Elite Award and 0 otherwise. The reference category is *Correctly Recognized High Performer*, which takes the value of 1 if the participant performed above average on their aptitude test and was randomly assigned the Elite Award and 0 otherwise. Model 2 introduces *Unfair*, which takes the value of 1 if the participant reported three or less when asked about the fairness of Elite Award allocation. Model 2 does not include controls because the level of fairness felt by a participant is likely related to these control variables. Robust standard errors are in parentheses. Significance levels: +p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.

Table 3. OLS Regressions of Recommend Elite by Evaluators' Recognition and Criteria (Study 3)

	Model 1	Model 2
Underrecognized	-0.069* (0.033)	-0.058+ (0.033)
Correctly Recognized Low Performer	-0.071 (0.048)	-0.083 (0.051)
Overrecognized	0.295*** (0.048)	0.330*** (0.052)
Criteria	-0.006 (0.033)	-0.006 (0.034)
Underrecognized × Criteria	-0.050 (0.046)	-0.051 (0.047)
Correctly Recognized Low Performer × Criteria	0.037 (0.054)	0.065 (0.057)
Overrecognized × Criteria	-0.100+ (0.055)	-0.157* (0.062)
Perceived Peer Performance	0.174*** (0.010)	0.179*** (0.011)
Time Spent on Peer Test	-0.000 (0.000)	-0.000 (0.000)
Number Marked Correct on Peer Test	0.060*** (0.008)	0.049*** (0.010)
Accuracy on Peer Test	0.020 (0.073)	0.054 (0.082)
Participant Test Number Correct	-0.021* (0.010)	-0.027* (0.011)
Participant Test Time Spent	0.000 (0.000)	0.000 (0.000)
Age	0.001 (0.001)	0.002 (0.001)
College Educated	0.018 (0.020)	0.029 (0.021)
Female	0.045* (0.019)	0.041* (0.020)
Democrat	-0.037+ (0.021)	-0.018 (0.022)
Republican	0.008 (0.028)	0.017 (0.029)
White	-0.050* (0.021)	-0.051* (0.022)
Full-Time Employee	0.000 (0.020)	-0.005 (0.021)
Constant	-0.533*** (0.110)	-0.518*** (0.115)
Observations	1,740	1,547
R-Squared Adj.	0.350	0.314

Note: Unit of analysis is the participant's evaluation of peer test. *Recommend Elite Award* takes the value of 1 if the participant recommended that the peer receive the Elite Award and 0 if they did not recommend it. *Criteria* takes the value of 1 if participant was randomly given explicit Elite Award criteria and 0 if not. *Underrecognized* takes the value of 1 if the participant performed above average on their aptitude test and was not randomly assigned the Elite Award and 0 otherwise. *Correctly Recognized Low Performer* takes the value of 1 if the participant performed below average on their aptitude test and was not randomly assigned the Elite Award and 0 otherwise. *Overrecognized* takes the value of 1 if the participant performed below average on their aptitude test and was randomly assigned the Elite Award and 0 otherwise. The reference category is *Correctly Recognized High Performer*, which takes the value of 1 if the participant performed above average on their aptitude test and was randomly assigned the Elite Award and 0 otherwise. Model 1 includes full sample of participants, and Model 2 limits the sample to participants who indicated that the peer answered fewer than seven questions correctly. Robust standard errors are in parentheses. Significance levels: +p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.