Johnson
Cornell
SC Johnson College of Business

Administrative Science Quarterly
1–44

DOI: 10.1177/00018392211010118
journals.sagepub.com/home/asq
SAGE

The Invisible Cage: Workers' Reactivity to Opaque Algorithmic Evaluations

Hatim A. Rahman[1]

Abstract

Existing research has shown that people experience third-party evaluations as a form of control because they try to align their behavior with evaluations' criteria to secure more favorable resources, recognition, and opportunities from external audiences. Much of this research has focused on evaluations with transparent criteria, but increasingly, algorithmic evaluation systems are not transparent. Drawing on over three years of interviews, archival data, and observations as a registered user on a labor platform, I studied how freelance workers contend with an opaque third-party evaluation algorithm—and with what consequences. My findings show the platform implemented an opaque evaluation algorithm to meaningfully differentiate between freelancers' rating scores. Freelancers experienced this evaluation as a form of control but could not align their actions with its criteria because they could not clearly identify those criteria. I found freelancers had divergent responses to this situation: some experimented with ways to improve their rating scores, and others constrained their activity on the platform. Their reactivity differed based not only on their general success on the platform—whether they were high or low performers—but also on how much they depended on the platform for work and whether they experienced setbacks in the form of decreased evaluation scores. These workers experienced what I call an "invisible cage": a form of control in which the criteria for success and changes to those criteria are unpredictable. For gig workers who rely on labor platforms, this form of control increasingly determines their access to clients and projects while undermining their ability to understand and respond to factors that determine their success.

Keywords: algorithm, labor platform, invisible cage, reactivity, control, opaque, evaluations, gig work

Third-party evaluations are a central feature of today's societal and organizational landscape (Lamont, 2012; Sharkey and Bromley, 2015; Espeland and Sauder, 2016). Studies have shown that third-party rating evaluations of actors

[1] Northwestern University, Kellogg School of Management

such as doctors (RateMDs), professors (RateMyProfessors), hotels (TripAdvisor), restaurants (Yelp), corporations (*Forbes*), and universities (*U.S. News & World Report*) provide a sense of transparency and accountability for external audiences (Strathern, 2000; Power, 2010; Orlikowski and Scott, 2014). Audiences also use third-party evaluations to form their perceptions and make decisions about the evaluated actor (Karpik, 2010). As a result, these evaluations influence the resources, recognition, and opportunities actors receive from external audiences (Pope, 2009; Brandtner, 2017). As the prevalence and influence of third-party evaluation systems have increased, researchers have examined how actors subject to such systems react to them (Jin and Leslie, 2003; Espeland and Sauder, 2007; Chatterji and Toffel, 2010). For example, because university admissions, funding, and recognition are influenced by third-party evaluations (e.g., *U.S. News & World Report*), university administrators and faculty pay close attention to the criteria these evaluations use, such as career placement statistics, and change their behavior to better align with them (Sauder and Espeland, 2009; Espeland and Sauder, 2016).

Consequently, while prior work has shown that third-party evaluations often provide transparency and accountability for external audiences, it has also suggested that actors subject to third-party evaluations experience them as a form of control (Espeland and Sauder, 2016; Brandtner, 2017; Kornberger, Pflueger, and Mouritsen, 2017). Because third-party evaluations can influence actors' ability to secure resources and recognition from their primary audiences, actors will likely internalize evaluations' criteria and change their behavior to conform to those standards (Sauder and Espeland, 2009; Masum and Tovey, 2011). Scholars label the phenomenon of people changing their perceptions and behavior in response to being evaluated as "reactivity" (Espeland and Sauder, 2007).

Technological advancements have expanded the use of third-party evaluations to new areas of work and organizing, raising new questions in this domain (Fourcade and Healy, 2016; Kuhn and Maleki, 2017; Cameron, 2021). Nowhere is this more evident than in the rise of labor platforms and their use of third-party evaluations to assess workers. While several types of platforms exist (Davis, 2016; Sundararajan, 2016), those most relevant to this study are labor platforms facilitating gig work, such as Upwork, TopCoder, and TaskRabbit. They provide a digital infrastructure to connect clients with freelance job seekers for relatively short-term projects. Labor platforms have attracted increased attention from work and organizational scholars because they differ from intermediaries and exchange systems previously studied (Vallas and Schor, 2020; Rahman and Valentine, 2021; Stark and Pais, 2021), particularly in their use of evaluations (Kornberger, Pflueger, and Mouritsen, 2017).

Unlike previously studied settings, in which third-party evaluation criteria are relatively transparent to those being evaluated, in labor platforms these criteria are often opaque to workers. This opacity makes it easier for platforms and clients to differentiate among workers by using their evaluation scores, because it is more difficult for workers to game and inflate the evaluation system than in traditional settings (Filippas, Horton, and Golden, 2019; Garg and Johari, 2020). Platforms' use of opacity in worker evaluations raises an underexplored question: how do opaque third-party evaluations influence workers' reactivity, and what mechanisms contribute to this form of reactivity?

While existing organizational research (Proctor, 2008; Briscoe and Murphy, 2012; Burrell, 2016) has broadly suggested that opacity will make it more difficult for workers to understand evaluation criteria, we lack grounded theory examining how workers contend with such opacity—and with what consequences.

To address this gap, I studied one of the largest labor platforms focused on higher-level project work, such as software engineering, design, and data analytics. The platform implemented an opaque algorithmic rating evaluation to better differentiate which freelancers should be visible to clients and to prevent freelancers from gaming their scores. Freelancers tried but generally failed to understand the evaluation's inputs, processing, and output, which led them to experience the opaque evaluation as a system of control characterized by unpredictable updates, fluctuating criteria, and lack of opportunities to improve their scores. These experiences were especially frustrating because the opacity contrasted with workers' expectations of employee-evaluation systems based on previous experiences in traditional organizations, where such systems' main purpose is to help workers improve (Cappelli and Conyon, 2018).

I observed that freelancers responded to evaluation opacity with two types of reactivity: they either tested different tactics to increase their scores, such as working on various types of projects and with different contract lengths, or they tried to preserve their scores by limiting their engagement with the platform, such as by working with platform-based clients outside of the platform and not working with new clients. This was the case both for workers with higher and lower scores on the platform. Two mechanisms determined their type of reactivity: the extent to which freelancers depended on the platform for work and income and whether they experienced decreases in their evaluation scores (regardless of whether those scores started out higher or lower). My findings support the argument that opaque third-party evaluations can create an ''invisible cage'' for workers, because they experience such evaluations as a form of control and yet cannot decipher or learn from the criteria for success and changes to those criteria.

THIRD-PARTY EVALUATIONS, REACTIVITY, AND OPACITY

Convergent Reactivity toward Third-Party Evaluations

Third-party evaluations do not simply act as objective measures of quality and performance for external audiences (Elsbach and Kramer, 1996; Espeland and Sauder, 2007; Chatterji and Toffel, 2010). Rather, they engender strong reactions from those who are being evaluated, because external audiences use these evaluations to form their perceptions and make decisions about the evaluated entity (Chatterji and Toffel, 2010; Espeland and Sauder, 2016; Brandtner, 2017). Espeland and Sauder (2007: 1) applied the label *reactivity* to ''the idea that people change their behavior in reaction to being evaluated, observed, or measured.'' Researchers have also shown that third-party evaluations can be experienced as a form of control even by those who are not directly evaluated. When Sharkey and Bromley (2015) examined firms' reactivity to third-party environmental rating evaluations that measure firms' toxic pollutant emissions, they found that even unrated firms try to reduce their emissions in ways similar to rated organizations in hopes of gaining legitimacy and external audiences' recognition.

The pursuits of legitimacy and audience recognition thus are mechanisms that inspire *convergent reactivity*: the tendency for individuals and entities to align their behaviors with the criteria used for third-party evaluations, which leads to similarities in behavioral changes among the various people and entities being evaluated. Scholars have identified two other mechanisms that also contribute to convergent reactivity in response to third-party evaluations. The first is self-fulfilling prophecy. Espeland and Sauder (2016) showed, for example, that universities tie their budget allocations to third-party evaluations, thereby encouraging faculty and administrators to align their behavior with the evaluation criteria. Second is *commensuration*—reducing disparate information into a single metric that is easy to use for judgment, categorization, and comparison (Espeland and Stevens, 1998; Bermiss, Zajac, and King, 2014). Commensuration directs attention to a narrower set of actions that are measured by third-party evaluations and downplays other actions (Espeland, 2015).

An important assumption in this literature is that people can change their behavior because evaluation criteria are relatively transparent and change infrequently (Brandtner, 2017). The stability and clarity of third-party evaluation criteria make it easier to determine and change one's behavior to align with the evaluation's expectations. In some cases the convergence in behavior is unproblematic, such as when universities take similar actions to increase their job placement percentages. But when differentiation is desired, behavioral convergence could undermine the legitimacy of a third-party evaluation (Filippas, Horton, and Golden, 2019). This issue has become particularly salient on platforms, including labor platforms (Garg and Johari, 2020).

Labor Platforms' Increased Use of Opaque Third-Party Evaluations

Labor platforms theoretically differ from traditional organizations, intermediaries, and market settings in important ways (Vallas and Schor, 2020; Lei, 2021; Stark and Pais, 2021), including in their use of rating evaluations (Kornberger, Pflueger, and Mouritsen, 2017). Emerging work on platforms' use of evaluations has shown that it differs from previously studied third-party evaluations in several ways: the public visibility of the evaluations, the centrality of platforms' use of evaluation scores to control workers' opportunities; the frequency with which evaluation scores can change; and the opacity surrounding how evaluations operate (Leung, 2014; Pallais, 2014; Burrell, 2016; Tadelis, 2016; Kuhn and Maleki, 2017).

Platforms use algorithms in real time to collect rating evaluations that clients give to freelancers after a project is complete (Rosenblat and Stark, 2016). Algorithms can aggregate a freelancer's rating scores across projects into a single score, which is publicly displayed on their profile, and use these aggregated scores to control freelancers' visibility and opportunities (Pallais, 2014), making it easier or harder for clients to find their profiles on a platform depending on their scores (Chan and Wang, 2018). Compared with previously studied third-party evaluations, these scores are updated more frequently given the number of projects or interactions users have on platforms (Horton, 2010; Rosenblat and Stark, 2016; Shapiro, 2017; Gray and Suri, 2019).

While many platforms initially used transparent evaluation criteria, they found that such systems can easily be gamed such that nearly all participants appear to have identical, perfect rating scores (Filippas, Horton, and Golden, 2019; Garg and Johari, 2020). Having many freelancers with similar rating evaluation scores can be problematic, because the low variation in scores makes it difficult for the platform and clients to differentiate freelancers (Akerlof, 1970; Tadelis, 2016). In other words, when evaluation scores can be gamed, they become a less effective method for matching freelancers with clients (Filippas, Horton, and Golden, 2019).

One reason platforms have been unable to control gaming behavior is that, almost uniformly, platforms claim to have no legal relationship with the workers listed there (Aloisi, 2015; Dubal, 2017). Instead, workers are treated as independent contractors or freelancers for the clients who hire them through the platform (Cherry, 2015). This classification has met intense social, legal, and political pushback, in part because many platforms control aspects of how workers must act when working for them, but workers can technically set their own schedules, receive remuneration on a per-project basis, and use their own tools, resources, or equipment to complete the work (Shapiro, 2017; Vallas and Schor, 2020; Cameron, 2021). Although legal cases questioning the relationship between platforms and workers are increasingly presented in court, there is still no clear resolution regarding how to classify the relationship between platforms and workers using them. In one prominent case brought by Lyft drivers against the platform, the presiding judge remarked that the court was "handed a square peg and asked to choose between two round holes. The test the . . . courts have developed over the 20th Century for classifying workers isn't very helpful in addressing this 21st Century problem. Some factors point in one direction, some point in the other, and some are ambiguous."[1]

Even in recent cases that have favored classifying platform workers as employees (Scheiber, 2018; Conger and Scheiber, 2019), the rulings have been specific to each case and lacking generalizability, sometimes resulting in unintended consequences.[2] My goal is not to settle the debate on the nature of the relationship between platforms and workers—a rapidly evolving dynamic—but rather to highlight that contemporary platforms theoretically (and legally) cannot subject workers to control measures of the past, including those illuminated by previous research (Vallas and Schor, 2020; Stark and Pais, 2021). This is in part because explicitly exerting control over workers could jeopardize the employment relationship between platforms and workers. In this context, a growing number of scholars have highlighted how labor platforms use opacity to exert influence over workers and prevent gaming behavior. In fact, because of the large power asymmetries favoring platforms and the clients who evaluate workers, many scholars have suggested that platforms have further tightened the iron cage or created a "post-panoptic" environment for workers, because workers are now exposed to even more restrictive expectations embedded in platforms' rules and evaluations (Fourcade and Healy, 2016; Finn, 2017; Curchod et al., 2019).

[1] https://blog.ericgoldman.org/archives/2015/03/court-says-uber-and-lyft-drivers-may-be-employees.htm, last accessed March 23, 2021.
[2] For example, see California Assembly Bill 5, September 2019 and Pofeldt, 2019 for coverage of the unintended consequences.

Increasing Focus on Opacity

Though scholars acknowledge the presence of opaque practices, research in work and organizational studies examining the use of opacity and its impact on workers is sparse (Proctor, 2008; Pasquale, 2015) for at least two reasons. First, organizations typically implement opaque systems and practices precisely to avoid close inspection of and attention to related processes (Oreskes and Conway, 2011; Briscoe and Murphy, 2012; Pasquale, 2015). As Briscoe and Murphy (2012: 554–555) noted, "Opaque practices are those for which observers have difficulty identifying key characteristics, including what is being done, to what degree it is being done, when the effects will transpire, and exactly who will have caused them." As a result, workers and other audiences may not even be aware of the opaque practices in place.[3] Second, and related, researchers have focused overwhelmingly on organizational phenomena broadly—and control tactics specifically—that are visible and thus easier to document (Proctor, 2008). So we have an incomplete theoretical framework about how intentionally opaque practices operate and influence behavior.

Despite the difficulty of researching opaque practices, many scholars have highlighted the need for greater scrutiny of such practices as we gain awareness of their impact on workers and society (Proctor, 2008; Briscoe and Murphy, 2012; Pasquale, 2015; Zuboff, 2019). The extent to which organizations use opacity in their operations has prompted the suggestion that we have become a "black box society" undergirded by "surveillance capitalism" in which organizations leverage both power- and information-related asymmetries to collect, analyze, and influence people in ways we are just beginning to understand (Pasquale, 2015; Zuboff, 2019). These metaphors starkly contrast the iron cage and "visible hand" images that have dominated our view of organizations and markets (Chandler, 1977; Weber, 1978; Barker, 1993). To build theory in this domain, I conducted an inductive analysis of freelance workers' reactivity when being subject to an opaque third-party evaluation algorithm in a labor platform setting.

METHODS

Research Setting

I examined workers' reactivity to opaque third-party rating evaluations they received on TalentFinder (a pseudonym), one of the largest organizations providing an online labor platform. From inception, TalentFinder positioned itself as an intermediary providing the technology infrastructure to match individual clients or organizations with freelancers, earning a percentage fee from the money clients paid to freelancers (Horton, 2010; Kittur et al., 2013; Kuhn and Maleki, 2017). Intermediaries matching labor with capital are not new (Autor, 2009; Bonet, Cappelli, and Hamori, 2013). Unlike previous intermediaries, however, TalentFinder and other emerging organizations provide a digital, centralized platform that could be used by almost anyone in the world. Recent estimates indicate that labor platforms are among the fastest growing markets (Chan and Wang, 2018).

[3] When workers do become aware of opaque practices, they typically undertake considerable efforts, such as whistleblowing or lawsuits, to shed light on the systems' existence (Proctor, 2012; Butler, Serra, and Spagnolo, 2020).

In 2015, TalentFinder had over 12 million registered freelancers and 5 million clients spread across 100 countries. Clients posted 3 million projects, generating over $1 billion of work. TalentFinder differentiated itself from other labor platforms by marketing itself as a platform on which clients could hire freelancers to complete administrative as well as complex, high-skilled work such as software, mobile application, and web development, along with design, animation, sales, and marketing.

TalentFinder's user agreements and arm's-length relationship with freelancers and clients. TalentFinder established its relationship with freelancers (and clients) when they signed up to use the platform. When freelancers registered, they were presented with links to TalentFinder's terms of service, user agreement, and privacy policy. Freelancers were not required to click on the links to indicate that they, at the very least, knew these agreements existed; instead, they could simply check a text box indicating they ''understand'' and ''agree'' to the respective policies. By selecting that checkbox and using the platform, users accepted that TalentFinder could make any changes to these agreements without explicitly notifying them, as stated in the user agreement:

> Subject to the conditions set forth herein, TalentFinder may, in its sole discretion, amend this Agreement and the other Terms of Service at any time by posting a revised version on the Site. . . . You understand that by using the site or site services after the effective date, you agree to be bound by the terms of service.

The agreement also included freelancers' ''consent'' for implementation and modification of the evaluation system TalentFinder used. Online Appendix A (http://journals.sagepub.com/doi/suppl/10.1177/00018392211010118) highlights portions of the user agreement related to TalentFinder implementing its evaluation system and establishing itself as a third-party evaluator in relation to freelancers. TalentFinder's user agreement further stated:

> TalentFinder merely makes the Site Services available to enable Freelancers to identify and determine the suitability of Clients for themselves and to enable Clients to identify and determine the suitability of Freelancers for themselves. TalentFinder does not, in any way, supervise, direct, or control Freelancer or Freelancer's work.

This provision was key in establishing TalentFinder as a third party to freelancers' and clients' interactions. In practice, however, TalentFinder did much more than making ''Site Services available.'' For example, it vetted and verified freelancers' legal identities upon registration, maintained a detailed record of freelancers' history on the platform, and used rating evaluations as a measure of freelancers' reputations to organize search results and suggest freelancers to clients. TalentFinder also provided payment-processing services and offered clients monitoring tools and other features they could use while working with a freelancer. But TalentFinder left it up to clients to determine who to hire, when and how much to pay, the type of contract to use, how long to work with the freelancer, and other conditions specific to each project. Freelancers, for their part, were theoretically free to choose which jobs to accept, negotiate their own wages, and determine when they wanted to work on the platform.

Evaluations on TalentFinder. During the study period, TalentFinder initially provided freelancers with two types of transparent five-star rating evaluation scores on the platform: project scores and overall scores. At the end of each project, a client could rate a freelancer on a 1 to 5 scale on dimensions including "Skills," "Quality of Work," "Availability," "Adherence to Schedule," "Communication," and "Cooperation." TalentFinder aggregated clients' scores on each dimension to provide an overall evaluation score for each project. The platform then generated an overall score for each freelancer based on their project evaluation scores, which were weighted by the dollar value of each project, with higher-valued projects weighing more. In search results, freelancers' profiles included their overall score out of five stars.

TalentFinder prominently used this rating evaluation system to sort, organize, and showcase different freelancers to potential clients. But the system had several limitations, especially rampant rating inflation. In the year before TalentFinder introduced a change to its evaluation, more than 90 percent of freelancers on the platform had at least a 4.00 out of 5.00 rating evaluation score, and 80 percent had close to a perfect rating. TalentFinder viewed such inflation as problematic because it rendered use of rating evaluation scores less effective as a matching mechanism. When everyone had similar evaluation scores, it was harder to use the scores to differentiate among workers and suggest freelancers in clients' search results on the basis of their past performance. Because clients and workers did not have access to or views of the larger dynamics on the platform, they were not aware that rating inflation existed and thus did not recognize it as a problem. Clients used the rating evaluation scores to search for freelancers and then used additional factors, such as looking at their past work experience, to decide which freelancers to interview and work with.

To address score inflation, in 2015, TalentFinder moved to a new algorithm to calculate and represent freelancers' reputation and performance based on calculation of an evaluation score on a 100-point scale. After its implementation, approximately 5 percent of freelancers were given what was considered a top score of 90 or above. When introducing the new system, TalentFinder did not explain how scores were calculated. The five-star overall evaluation was removed from each freelancer's profile and replaced with a percentage (i.e., "85%") indicating the new score. This score was prominently displayed and one of the only indicators differentiating freelancers in search results; see Online Appendix B. Further, clients could filter their search results by this new score. If clients were interested in seeing freelancers' five-star project evaluation scores, they now had to click on a freelancer's profile and scroll down to see the previous projects completed and how clients evaluated them.

Data Collection

The initial motivation for this study was to understand freelancers' experience of working on the platform, broadly, and their specific experience with the algorithmic evaluation. Similar to previous inductive studies examining the nature of freelancers' experiences (e.g., Evans, Kunda, and Barley, 2004), I sought to build theory by collecting data providing insight into their broader experiences. Based on emerging studies on labor platforms (Rosenblat, 2018; Cameron, 2021; Rahman and Valentine, 2021) and my own experience using the platform, I collected data from three primary sources: interviews, archival sources, and my observations as a registered client and freelancer. I collected data from these sources iteratively, primarily between 2015 and 2018, with each source providing a different lens into the focal phenomenon, which was helpful in distinguishing findings that were fleeting from those that were consistent across the data and experienced more generally by platform users. Collecting data from multiple sources helped me to achieve greater consistency and completeness and enabled corroboration of findings from each data source.

Interviews. I began interviews for this study in 2017, 2.5 years after TalentFinder introduced the new algorithmic evaluation. Rather than using the random sampling approach common in deductive studies, I employed a theoretical sampling approach to collect interview data. Such approaches, common for inductive studies, involve purposely selecting informants to maximize access to data on theoretically relevant categories and phenomena (e.g., Li and Piezunka, 2020). In this case, TalentFinder publicly stated that "any score at 90% or above is excellent." Because I was most interested in understanding freelancers' experience with the algorithmic evaluation, I based my sampling strategy on their evaluation scores. I interviewed freelancers who had what TalentFinder considered a high score ($\geq$ 90 percent), freelancers with scores below that threshold, and workers new to the platform who did not yet have evaluation scores. I used my client account to create a job posting and recruit freelancers for interviews. The job posting, titled "Understanding Freelancing Experience," invited freelancers to share their experiences working on the platform; see Online Appendix C for the job description I used.[4]

I conducted interviews and kept the job post active until I reached "theoretical saturation" (Glaser and Strauss, 1999) and was no longer able to glean new insights related to workers' experiences with the algorithmic evaluation—that is, until the details provided by new interviewees overlapped with and corroborated the experiences of previously interviewed freelancers. This approach yielded a total of 80 semi-structured freelancer interviews: 36 with freelancers who had an evaluation score at or above 90 percent and were considered high-performing by TalentFinder; 26 with freelancers with an evaluation score below 90 percent (ranging from 33 to 89 percent), who were considered low-performing freelancers by the platform; and 18 with freelancers who did not yet have evaluation scores.[5] I interviewed freelancers who were new to the platform (with no evaluation scores) to understand what factors they believed were important to succeeding on TalentFinder, the extent to which they were aware of the evaluation algorithm, and whether they believed it would contribute to their success. Of the 62 interviewees who had received an evaluation

[4] Online Appendix F includes additional details about how I set up and recruited participants for interviews.

[5] To augment the study's findings, I conducted 21 of the interviews 4.5 years (2020) after the introduction of the algorithmic evaluation, using the same sampling protocol and questions. Online Appendix G discusses these data in more detail.

score, 37 had experienced both the prior five-star evaluation system and the current evaluation system.

All interviews lasted between 30 and 60 minutes. Because interviewees were located worldwide, I conducted interviews using each freelancer's choice of voice/phone communication (e.g., Skype). The semi-structured interviews focused on their overall experience working on the labor platform. While talking about their experiences on the platform in general and with the algorithmic evaluation, some freelancers shared their screen to show me their profile and previous conversations with other freelancers. Online Appendix D lists the information I collected for each freelancer, including their evaluation score at the time of the interview, how long they had been on TalentFinder, and whether they experienced the previous evaluation system, along with their listed job specialty, country, gender, and preferred hourly wage. Interviewees represented 17 job specialties and 19 countries, and the sample was relatively balanced between male (42) and female (38) workers.

Although I aimed mainly to understand how the algorithmic evaluation affected freelancers' work on the platform, I also wanted to understand clients' awareness of their impact on freelancers' evaluation scores. But clients were much more difficult to recruit and interview. Registered freelancers could see a client's full profile only after applying for their job, and even then, TalentFinder withheld their contact information. Given the challenges of this approach, I instead relied on my freelancer informants to suggest clients as potential interviewees; 18 clients agreed to be interviewed. I used client interviews to understand their perspectives on providing freelancers feedback on the platform and to triangulate what freelancers said about their experiences working with clients. Online Appendix E lists the information I collected from clients, including how many freelancers they had hired, years on TalentFinder, location, whether they indicated they were aware of how their actions affected freelancers' evaluation scores (beyond the public five-star evaluation), and whether a client used what TalentFinder considered a top score (≥ 90) as a filtering criterion to find freelancers.

Archival data. TalentFinder provided freelancers a publicly available discussion board to post questions and answers related to work and non-work-related issues encountered on the platform. In interviews, freelancers told me that they relied on the discussion boards, so I collected community discussion board messages related to the algorithm, focusing on the most popular threads. This approach yielded 2,053 TalentFinder community discussion board messages from 28 threads related to the algorithmic evaluation, dating back to the algorithm's initial introduction. I also periodically checked the discussion boards after I collected my interview data, looking for the emergence of new themes. In addition, I collected all of TalentFinder's public posts related to the algorithm, which included blog posts, help articles, and moderators' comments on discussion boards.

Registered user. To collect firsthand data, I became both a registered client and a freelancer on TalentFinder. Even registering to use the platform proved fruitful for collecting data, as I gained detailed information about the terms of service to which freelancers and clients were bound, the types of information

TalentFinder requested from each type of user, and the information TalentFinder provided to new workers and clients. Once registered, I received e-mails with tips and suggestions and had access to useful resources and privileges. I kept a detailed account of the e-mails, notifications, and platform changes I observed through my accounts. On the TalentFinder platform, I posted jobs for freelancers, paid them for their time and effort, and left feedback and evaluations after projects ended.

As a registered client, I gained firsthand insight into how TalentFinder communicated with clients and learned what it was like to search for and hire freelancers and use TalentFinder's evaluation system. As a registered freelancer, I collected information on how TalentFinder formally communicated during registration and when the organization made changes to its platform. I also gained basic knowledge about using the platform as a freelancer. After speaking with freelancers and based on my own ethical considerations, I decided not to apply for jobs because there was a shortage of project supply compared with available freelancers. Table 1 provides a summary of my data collection sources, how I collected data from each source, and how I used the data sources in my analysis.

Data Analysis

I followed an inductive approach in analyzing these data sources, iterating between data from the interviews, archival data, and my own experiences (Glaser and Strauss, 1999; Charmaz, 2006). When I open coded the interview data, freelancers' frustration with the algorithmic evaluation stood out in almost every interview. I was initially surprised by this sentiment because more than two years had passed since the introduction of the new evaluation. To see if this initial finding was particular to my interview data, I open coded the data collected from the discussion boards, help articles, blog posts, and any communication I received and collected from TalentFinder as a registered client and freelancer. The archival data were especially useful because I tracked a broader range of freelancers' experience with the algorithmic evaluation from its commencement to my data collection period and even after I collected my interview data. I found that several themes recurred in my interviews and the archival data, reflecting key aspects of the freelancing experience on TalentFinder.

Specifically, open coding of the archival data showed that freelancers were unaware that the algorithmic evaluation was being introduced and did not know what factors contributed to the new scoring system. Surprisingly, this confusion did not subside over time; freelancers continued to share their experiences with the algorithmic evaluation and theories about why their scores went up or down, often with differing accounts. To assess whether this initial analysis triangulated with other data sources, I systematically went through each help article and blog post related to the algorithmic evaluation and any communication I received from TalentFinder. I found that TalentFinder never clearly specified the criteria, their weighting for the algorithmic evaluation, or how often the evaluation score changed. Thus my initial coding of all three data sources highlighted that the algorithmic evaluation was highly opaque to freelancers.

To develop a more refined analysis, I coded the data to identify the specific components of freelancers' experiences that described their understanding of

Table 1. Summary of Data Collection

Type	Description	Sampling Strategy	Use in Analysis
Interviews	98 semi-structured interviews: ○ 80 freelancer interviews: • 36 had a score $\geq$ 90 • 26 had a score between 33 and 89 • 18 were new to the platform and were not provided a score ○ 18 client interviews	Freelancers: Theoretically sampled freelancers with a broad range of evaluation scores, particularly those whom TalentFinder considered high performers ($\geq$ 90), those below a high score, and freelancers new to the platform Clients: Interviewed clients referred by freelancers	Understand freelancers' experience with the algorithmic evaluation, broadly, and how their experiences differed Examine whether differences in freelancers' experiences and reactivity were based on their evaluation or other factors such as their specialty or location Identify mechanisms contributing to variation within and between high- and low-performing freelancers' reactivity Understand how clients perceived their role in providing feedback to freelancers and how they used the evaluation score to select freelancers
Archival data	Discussion board: Collected 2,053 freelancer messages from 28 threads posted on TalentFinder's discussion board TalentFinder communication: TalentFinder's announcements, blog posts, help articles, e-mails, and replies on the discussion boards	Discussion board: Selected most popular discussion board threads and posts (measured by number of freelancer messages on the thread) related to the algorithmic evaluation TalentFinder communication: Collected all TalentFinder communication that was publicly available	Explore the extent to which themes discussed by a broader range of freelancers on the discussion boards overlapped with themes from interviews Analyze whether the experiences and impact of opaque control changed over time Understand the online platform organization's public explanations, goals, and strategy related to implementing and maintaining an opaque algorithm Identify conditions enabling online platform organizations to implement opaque control tactics
Registered user	Collected all firsthand information provided to registered users, including e-mails and notifications, user agreement, terms of service, and data collected by the platform about users when using the platform	Spent four years registered as both a client and freelancer to experience the platform from both types of accounts	Gain firsthand account of registering, using, and experiencing changes made to the platform Collect TalentFinder's announcements and communication sent only to registered users Compare how TalentFinder's information sent to freelancers differed from that sent to clients Triangulate insights gleaned from interviews and archival data

the algorithmic evaluation and how it affected their actions. This step helped me identify which experiences were shared by all freelancers and where there was variation in freelancers' experiences. For freelancers' shared experiences, I looked for themes in the data that occurred regardless of differences in their experiences or performance on the platform. These themes included difficulty learning from the opaque evaluation and a feeling of paranoia about the algorithmic evaluation. I coded data in the ''difficulty learning from opaque evaluation'' category when freelancers expressed that changes in their evaluation score did not provide information that helped them understand how to improve their work performance or interactions with clients. The ''paranoia'' category was first inspired by freelancers' emic account of the impact the algorithmic evaluation had on them. In further coding I observed that several other freelancers described sustained uneasiness and heightened suspicion about how the algorithm operated and which actions affected their scores. For example, in the ''paranoia'' category I included instances when freelancers mentioned they had read every available article about the algorithm but still believed that any action they took could negatively impact their evaluation score.

To analyze whether freelancers' understanding, experience, and reported consequences differed, I divided the interview data into three groups based on interviewee type: freelancers with scores 90 or above (high performers), below 90 (low performers), and no score yet assigned (new freelancers). I looked at similarities and differences in my data across these categories to assess, for example, whether freelancers could use the advantages that came with having higher scores to better navigate the opaque control tactics they encountered. While all freelancers had access to the same information and were equally subject to TalentFinder's evaluation process, I observed variation in how freelancers reported responding to the evaluation, but the variation did not just depend on whether they were high or low performers. I reorganized the data into groups based on their responses, which helped me identify variation both within and between high and low performers' responses to the evaluation. When freelancers related that they tested different tactics to change their scores, I coded that response as *experimental reactivity* because freelancers were never sure how these tactics would turn out. Other freelancers reported trying to limit their exposure to the algorithmic evaluation, and I coded these practices as *constrained reactivity*.

I next coded the data to determine whether there were differences in the reasons that freelancers expressed for engaging in experimental or constrained reactivity. This step helped identify *platform dependence* and *evaluation setbacks* as common mechanisms contributing to experimental and constrained reactivity. I coded freelancers as having high platform dependence if they stated that they mainly relied on TalentFinder for finding work and income and low platform dependence if they said they also secured project work elsewhere. In interviews and archival data, freelancers frequently mentioned whether their evaluation scores had increased, decreased, or remained the same. Because they mentioned evaluation score decreases as being upsetting, I coded instances in which freelancers experienced score decreases as evaluation setbacks. In the findings section I describe how these two mechanisms contributed to differences between and within high and low performers' reactivity.

To further analyze the consequences I found in the interview data, I systematically analyzed the archival data for the prevalence of these themes over time. I examined whether these themes were present within 6 months, 1.5 years, and 2.5 years after the introduction of the algorithmic evaluation. I chose these timelines to ensure I was observing experiences that were lasting rather than fleeting. As I tried to understand how TalentFinder was able to implement the algorithmic evaluation, which affected freelancers so deeply, my excavation of the platform's documents also provided insight into the public reasons that TalentFinder launched the algorithm and how the platform implemented and maintained its opacity.

To develop a richer theoretical account of my findings, I considered what they "are a case of" more broadly by frequently iterating between the findings and existing literature. This process suggested that my preliminary conclusions resonated closely with the literature on workers' reactivity to third-party evaluations. I noticed that relative to my findings, past studies described experiences with more transparent third-party evaluations in which workers more easily understood the norms and expectations for behavior. These studies did not adequately account for what I had observed, particularly the difficulty freelancers had understanding what actions to take on the platform even several years after the algorithm was implemented.

When revisiting my findings with the fuller context of my analysis in mind, I noticed some freelancers described that TalentFinder made their profiles ''invisible,'' which illuminated a contrast to previous findings describing control tactics through the ''iron cage'' metaphor. Metaphors provide a powerful tool for theory building in organizational theory (Weick, 1989), and using one for this study helped to clarify how my findings were theoretically similar to or different from current theories of reactivity to third-party evaluations, as well as to delineate boundary conditions specific to this study. The ''invisible cage'' metaphor that I explicate in the following sections helped me see how freelancers' experiences with the opaque third-party evaluation contrasted with our current understanding of experiences with comparatively transparent third-party evaluations.

FINDINGS: INSIDE THE INVISIBLE CAGE

Unannounced Introduction of Algorithmic Evaluation

Freelancers reported that TalentFinder introduced the algorithmic evaluation without advance notice or explanation of how it would change the way they were scored or presented in searches. In line with this, neither my freelancer nor client account received any notification about this change.[6] As a high-performing freelancer told me, ''TalentFinder is notoriously bad at announcing changes . . . you would think they would announce it in the announcements [section of the website], but they did not do that for the rating algorithm'' (FI-14).[7] Instead, the freelancer continued, ''one morning I woke up and could not log in'' to their account. This freelancer initially thought someone may have stolen their password and account information, but they later attributed the difficulty to the introduction of the algorithmic evaluation. Another high-performing freelancer said of the evaluation's debut, ''As always, TalentFinder just does whatever the hell it wants without giving any explanation . . . they dropped it on us all at once like a nuclear bomb'' (FI-30).

The main surprise for freelancers was not that TalentFinder introduced a new feature—it had rolled out new features before, such as how notifications are displayed on the website—but that it introduced such a significant feature, which affected workers' appearance in clients' search results, without notice or explanation. As a high-performing freelancer said about the importance of their evaluation score, ''it is our billboard, it is our PR megaphone, it is the front door to our shop. . . . I guard it with my life, literally'' (FI-19).

[6] Although my data collection and analysis focus on freelancers' experiences after the algorithmic evaluation was implemented, I had a freelancer and client account two years before the system's debut and kept all communication records from the platform. I was able to use these records to verify freelancers' accounts related to notifications received from the platform.

[7] Freelancer quotes from interviews are demarcated by their ID listed in Online Appendix D. A quote from an interview with freelancer 14, for example, would be cited in the text as ''FI-14'' to indicate ''Freelancer Interview #14.'' Because archival data presented in the findings represent quotes from unique freelancers, I did not include an identifying numbering system for these data. To maintain freelancers' confidentiality, I use third-person-plural pronouns instead of gendered pronouns.

In my data, the earliest source of recorded information about the new evaluation was the community discussion board forums, but that information did not explain the new rating score. In the forum, a TalentFinder representative stated that ''as of now we're not planning to add an explanation to the blue bar display [i.e., algorithmic evaluation score].'' At that point, the new algorithmic evaluation had not fully replaced the five-star rating; however, even when TalentFinder later decided to make the algorithmic evaluation the sole representation of performance, freelancers did not report receiving notification that this change was taking place or what it meant for them.[8]

For freelancers, TalentFinder's lack of notification about the evaluation's existence did not improve through time. Those who joined the platform after its launch received no official notification or communication about the evaluation system. Not a single freelancer I interviewed who joined TalentFinder after the new evaluation was implemented had heard from the platform about the algorithm, even after completing their first jobs. And even if new freelancers somehow found out about the algorithmic evaluation when they started on the platform, TalentFinder did not reveal when they would actually receive a score.[9] Regardless of whether a freelancer was new or had been part of the platform since its inception, TalentFinder was circumspect about the criteria and weighting that the algorithm used to calculate their scores.

Differentiating Freelancers and Preventing Gaming through Opaque Evaluation Inputs and Processing

It was not until TalentFinder's first blog post discussing the algorithmic evaluation, three months after it introduced the new system, that freelancers indicated they could gain any insight into the reason the platform implemented it. But the blog post was not shared directly with freelancers. Only freelancers who monitored TalentFinder's blogs and discussion boards or discovered these announcements by happenstance noticed this post, which explained:

> We introduced the algorithmic evaluation score . . . as a new, more complete reflection of your client satisfaction. Thus far, the algorithmic evaluation score has proven successful at helping great freelancers stand out to clients and land more projects, with twice the number of contracts going to the best freelancers in the marketplace than before . . .

Importantly, TalentFinder stated that part of the motivation not to fully reveal the algorithmic evaluation's inputs or weighting was to ensure freelancers could not game the system:

[8] Although my analysis and findings focus on freelancers' experiences, clients also did not report receiving notification about the introduction of the algorithmic evaluation; this was the case for both the clients I interviewed and my own client account.

[9] In regard to when new freelancers might get an evaluation score, TalentFinder stated in a blog post that ''More than 90% of freelancers have a score after 5 projects. Since not all projects and clients are equal, the length of time will vary. Nearly all freelancers have scores after completing eight projects. The projects need to be with at least three different clients and take place within a 24-month period.'' None of the new freelancers I interviewed knew this, nor did I receive any communication from TalentFinder about this blog post or the information it communicated when I registered as a freelancer.

> We don't reveal the exact calculation for your score. Doing so would make it easier for some users to artificially boost their scores. We need to maintain some privacy with this metric to ensure fairness and accuracy.

In additional help and support articles, TalentFinder listed other inputs that may have led to changes in a freelancer's score. In practice, however, TalentFinder guarded closely the algorithmic evaluation's inputs and associated weighting. For example, in a blog post, TalentFinder indicated that a rating from a client who ''repeatedly receives poor feedback from freelancers'' would not affect a freelancer's rating evaluation score, and missing feedback would be counted against freelancers only when it ''represents a significant portion'' of their completed projects. Notably, the platform did not define ''repeatedly'' or ''significant'' in these statements. Instead, in subsequent posts and answers, TalentFinder representatives mainly encouraged freelancers to focus on completing ''quality product'' for each client. As a TalentFinder moderator advised freelancers questioning the vague language related to the evaluation's inputs and processing:

> You can complete many short-term jobs or do some occasional projects on TalentFinder, you can work on a few long-term contracts or work with repeat clients on different contracts. Having long term or repeat clients counts positively in the score but not having them doesn't count negatively. What matters the most is professionally communicating with the clients, successfully completing contracts and delivering quality product.

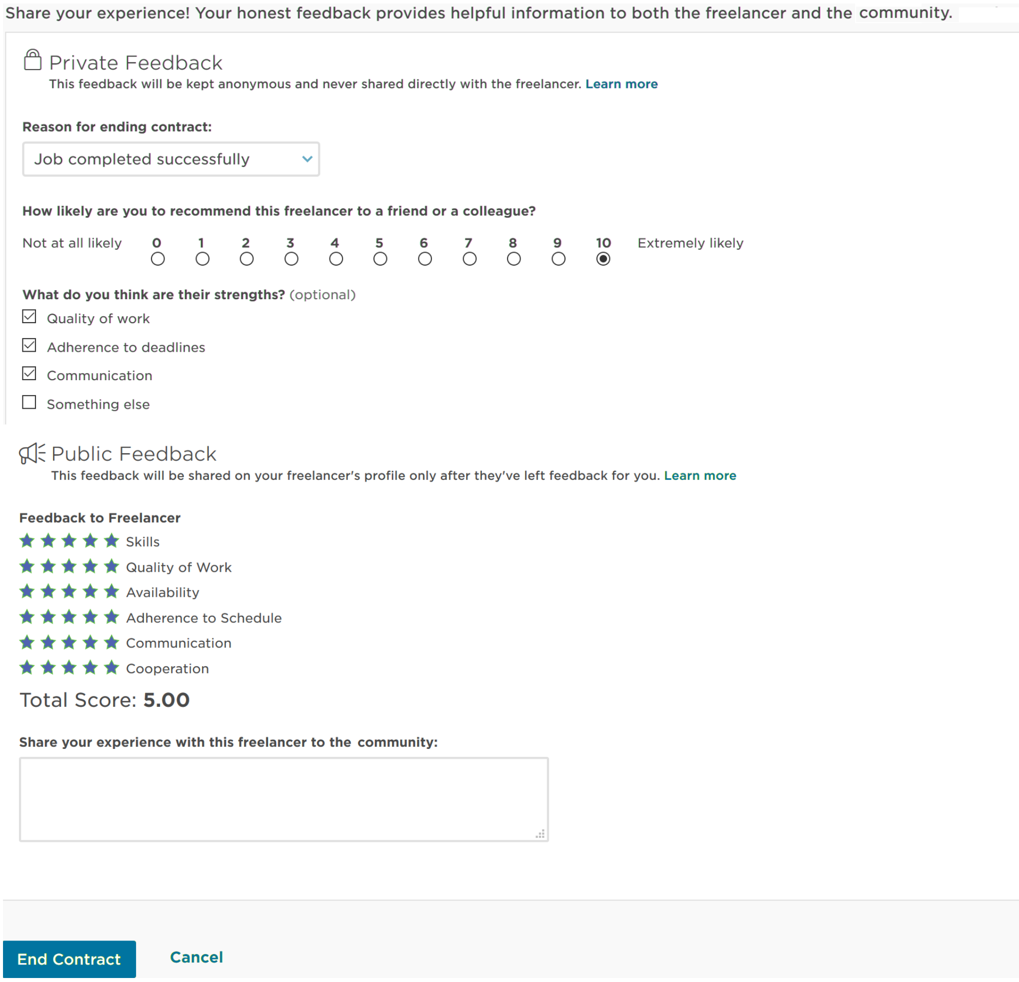
Opaque inputs: Private feedback. A big source of freelancers' frustration was the use of private client feedback. At each project's end, TalentFinder framed the overall feedback section by urging clients, ''Share your experience! Your honest feedback provides helpful information to both [the] freelancer and the TalentFinder community.'' Clients were then asked to rate on a scale of 1 (''not at all likely'') to 10 (''extremely likely''), ''How likely are you to recommend this freelancer to a friend or colleague?'' Before clients answered this question, they were told, ''This feedback will be kept anonymous and never shared directly with the freelancer.'' Figure 1 shows this feedback interface and the ease with which clients could fill out private feedback, without necessarily considering its impact on freelancers.[10]

Not only did freelancers express that they could not access clients' private feedback scores, but also they conveyed that they did not understand how private feedback affected their scores. Some freelancers believed that any rating below a 9 on the private feedback score lowered their evaluation score. One freelancer recounted their experience in detail on the discussion board:

> I am very disappointed in TalentFinder. I am a top-rated freelancer, I work hard to please my client. This client, I went through ENDLESS revisions that took me 2 months to complete just a single artwork with MEASLY pay. But I did it all for [my evaluation score]. Then 4 days ago, we ended the contract. It was the only contract I

[10] Screenshots have been slightly modified to preserve informants' and the platform's anonymity.

Figure 1. TalentFinder's Client Feedback Interface

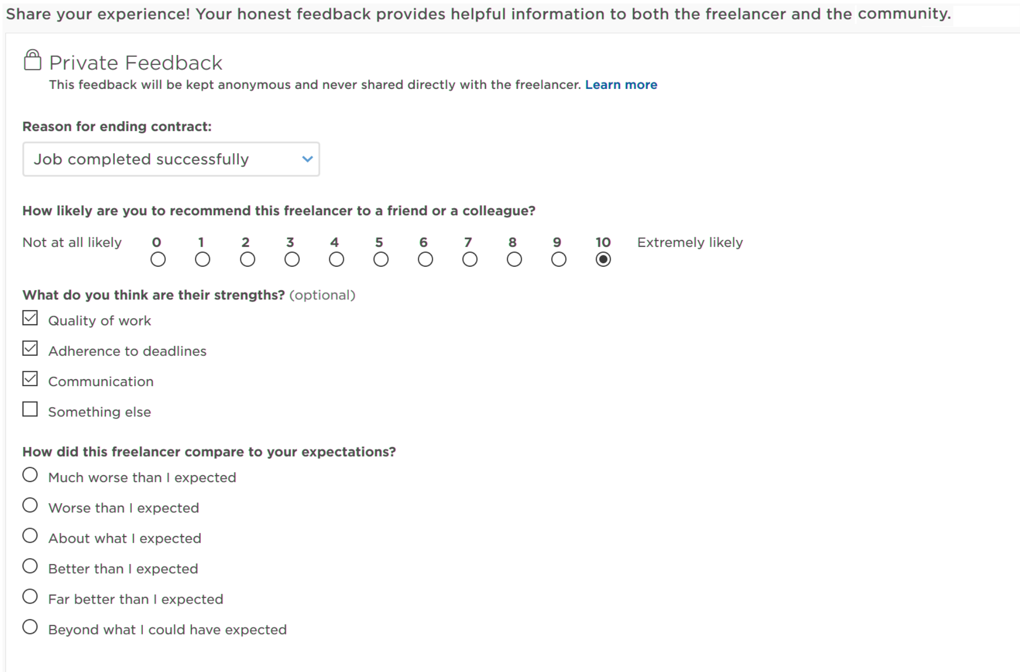


> ended in 2 weeks. And guess what? Today I woke up to 88% evaluation score. LOL, I was 98% before!!!

This freelancer stated a specific timeline for when the contract was ended and that no other projects had ended in that time frame. Freelancers explained that because it was unclear when and how the algorithm updated their evaluation scores, it was important to isolate whether any other actions or projects could have also contributed to score changes. In fact, the only reason this freelancer could hypothesize confidently about the source of the score decrease was because they personally reached out to ask what rating the client gave them in the private feedback section: ''The client told me she rated me '8' in private feedback.''

I also observed that the evaluation's inputs were subject to change without notice or explanation. After finishing an interview with a freelancer, I noticed that TalentFinder had added a new private feedback question without any

Figure 2. New Question Added to Client's Private Feedback Section



explanation or announcement for clients or freelancers. After ''How likely are you to recommend this freelancer to a friend or a colleague?'' the new question asked, ''How did this freelancer compare to your expectations?'' Figure 2 shows this question and possible responses, which were not numerical. None of the freelancers or clients I spoke to were aware of the question's impact on their evaluation score, and I could not find any communication from TalentFinder about it. Three months after I first saw it, the question was removed from the client feedback section without explanation or notification. Perhaps the question was added as an experiment, but freelancers and clients could not know how that question contributed to changes in their evaluation scores when it was present.

Opaque Inputs and Processing Contribute to Output Unpredictability

It was also difficult for freelancers to know when and how the algorithmic evaluation updated its inputs, parameters, and/or results. As an experienced, high-performing freelancer conveyed, their evaluation score ''fluctuated for no apparent reason'' (FI-3). Freelancers could not ascertain how their or clients' actions affected their scores in part because TalentFinder did not always reveal when a freelancer's evaluation score would update. When a contract was closed, a freelancer's score did not update right away, and TalentFinder did not notify freelancers when their scores did change. A freelancer shared that they ''did not get any notifications'' when their score changed and noticed that their score had dropped only when they looked at their public profile page, which was ''just shocking'' to see (FI-58).

Both decreases and increases in freelancers' scores were unpredictable. A high-performing freelancer remarked that after months of effort, their evaluation score finally increased, but ''The algorithmic evaluation is determining your fate, and you have no reason why it changes'' (FI-2). Another top-performing freelancer noted the algorithm's seemingly capricious output this way: ''Every closed contract has an impact on your score. What will that impact be? You will have to wait until the next calculation is made [to find out]'' (FI-24).

Difficulty in Learning from Output Unpredictability

The changing opaque expectations and unpredictable impact related to the algorithmic evaluation had consequences for how freelancers worked on TalentFinder. Even the most successful freelancers, whose scores never fell below the top score threshold, were affected. As a high-performing freelancer said, ''I understand why they went to the algorithmic evaluation with everyone on the platform [previously] having a five-star rating, but imagine receiving a grade in a class, but the grade was based on criteria you didn't know about . . . how are you supposed to improve'' (FI-15)? This comment reveals that freelancers' difficulty learning from the evaluation score was frustrating in part because they equated the score TalentFinder provided to evaluations they received in other contexts. In their experience and consistent with prior literature (Cappelli and Conyon, 2018), other evaluations, such as ''a grade in a class,'' provided input on how they could improve, but the algorithmic evaluation did not.

TalentFinder had not provided concrete guidelines with which freelancers could improve their scores. A freelancer on the discussion board wrote that ''there's nothing I can do to figure out how to improve my [evaluation score].'' The scores—and changes to them—provided little practical information for how to change behavior, which is generally the core purpose of evaluations in other work settings (Cappelli and Conyon, 2018).

Freelancers recounted that learning how to improve was also difficult because some clients did not realize that their scores in the private feedback section affected freelancers' scores. Unless a freelancer told a client about the importance of their private feedback to the algorithmic evaluation, none of the clients I interviewed understood the private ratings' impact. For instance, after explaining they give only positive feedback to freelancers, one of the most experienced clients I spoke to—who had hired over 150 people on TalentFinder—said, ''I do use the private feedback option, and it is not always in line with what I give for the public feedback'' (Client-I-18).

In addition, clients had differing interpretations of the scores they gave to freelancers. As a freelancer wrote on the discussion board:

> Many clients WILL NOT give a 9 or 10, even if they are thoroughly satisfied with the end product. It's who they are. And in their mind, giving an 8 or even a 7 is a perfectly ''good'' rating. They have no idea that it does serious damage to someone's ability to get future work on this platform.

Echoing this statement, a different freelancer on the discussion board added that ''a rating of less than 9 is viewed as unfavorable [by freelancers], and will impact the evaluation score. . . . I don't think clients are aware that they could destroy someone's future on [TalentFinder] by giving them an 8! Which is like a B or B+ and surely not a bad rating!'' Seeing their score decrease but not knowing what a client's private feedback was or whether private feedback contributed to the decrease underscored to this freelancer that there was no way to use their evaluation score as a means to improve. This was especially the case if a client's project communication and public feedback were positive or did not indicate how a freelancer should improve. As a different high-performing freelancer noted, ''If the client has never told me what was wrong, there is no way to fix it or make it better on future projects'' (FI-3).

Lack of Recourse

Even though TalentFinder gave freelancers the means to report a bad client or dispute a payment, there was no established way to appeal their score or file a complaint with TalentFinder about it. As a freelancer who experienced a decrease in their evaluation score said, ''The really annoying thing about this is that there is no recourse'' (FI-58). Some freelancers tried to contact TalentFinder's customer support, but to no avail. A high-performing freelancer told me, ''You learn very quickly that customer service is not helpful'' (FI-14). Another freelancer said that all customer service did ''is copy pasted messages that I can find on their FAQ'' (FI-42). Clients had similar concerns; once they completed their feedback, they had difficulty amending it. A freelancer on the discussion board described how a client accidentally completed the freelancer's private feedback with the wrong inputs but was not able to change it:

> The client [said] that her private feedback may have negatively impacted the freelancer's score (she [said she] hurriedly clicked on anything just to get past the window and rehire the freelancer!). She said she was unable to change her private feedback, even after reaching out to TalentFinder.

Difficulty Learning and Lack of Recourse Contribute to Paranoia

One of the most salient impacts of experiencing the difficulty in learning and lack of recourse in relation to their evaluation scores was that freelancers felt paranoid about the seemingly arbitrary algorithm. They felt sustained uneasiness and suspicion about how the algorithm operated and which actions led or could lead to changes in their scores. Even freelancers who maintained a perfect evaluation score for years felt this way. One such high-performing freelancer explained, ''I have read every article and piece of information I could find, and I still don't understand the darned thing . . . who knows, it could plummet for who knows what reasons'' (FI-6). Another freelancer on the discussion board elaborated:

> I haven't received a bad review, but I admit that reading the horror stories makes me paranoid that every slightly strange interaction with a client might lead to a bad review. [The algorithmic evaluation has] led to a whole feeling of unsettlement, nervousness, paranoia, low morale and suspicion among freelancers.

Freelancers' uneasiness was heightened when they could not pinpoint the reason their score was decreasing, as this high-performing freelancer expressed:

> I am one of those who are petrified with [the algorithmic evaluation]. My score fell from a 100 to 98 to 95 yesterday and you can check my profile and see my testimonials, I really work hard to make sure I deliver good work, I have been so worried that it literally is affecting my work. . . . I am all for rating, and I am ready to work harder but it still scares the living hell out of a lot of us. (FI-23)

Perhaps the most creative expression of how freelancers felt about the algorithmic evaluation's inner workings was conveyed in a tongue-in-cheek poem on the discussion board:

The [algorithm]
None can explain;
To attempt to decipher,
Is an effort in vain.
It's up, or it's down,
With no reason in sight;
Accept or do not,
You won't win the fight.
So work and work,
Leave the mysteries be;
To ponder [the algorithm],
Is a path to misery.

But freelancers could not simply ignore the algorithm. As a high-performing freelancer who maintained that rating for years said, "I have tried to tell myself not to care about the rating and just focus on the work," but the "most difficult part [of using TalentFinder] is always worrying about that stupid rating" (FI-21). This constant wariness had differing consequences for how high- and low-performing freelancers reacted to the opaque evaluation.

Experimental and Constrained Reactivity in Response to Opaque Evaluation

I observed variation within and between high and low performers' reactivity when they were subjected to TalentFinder's opaque algorithmic evaluation. Experimental reactivity was characterized by freelancers increasing their platform activity to try different tactics to increase their scores, such as testing different contract lengths or how contracts were closed. Constrained reactivity involved freelancers trying to limit their exposure to the evaluation; at times, this resulted in freelancers working with platform-based clients off-platform to ensure their transaction would not be recorded and hence not subjected to the evaluation.

Two mechanisms influenced whether high- and low-performing freelancers engaged in experimental or constrained reactivity: *platform dependence*, the extent to which they depended on the platform to find work and income, and *evaluation setbacks*, freelancers' experience of decreases in their evaluation scores. As I show next, high-performing freelancers who had high platform dependence diverged in whether they enacted experimental or constrained reactivity based on whether they experienced a decrease in their evaluation score: those who experienced a decrease enacted experimental reactivity, while those who did not enacted constrained reactivity. This pathway varied from other high-performing freelancers who had low dependence on the platform, as well as low-performing freelancers: these groups followed a pathway leading to experimental or constrained reactivity regardless of their experience with decreases in their evaluation score.

Figure 3 provides an overview of the variation within and between high and low performers' reactivity pathways and the mechanisms contributing to this variation. Table 2 shows the number and percentage of workers I interviewed who reported how the evaluation affected them on TalentFinder. The first two rows show the effect of consequences all freelancers experienced: difficulty learning and paranoia. The next two rows show the frequency of experimental and constrained reactivity, respectively.[11]

Experimental Reactivity: When High Performers with High Platform Dependence Have Evaluation Setbacks

Because TalentFinder offered freelancers access to clients around the world, many high performers depended on it to earn their income: ''I rely on TalentFinder to find work . . . it is easier than having to travel [to find work]'' (FI-29). When high performers who depended on the platform experienced evaluation setbacks, they rarely remained idle, even if the score decreases were minimal; as one freelancer said, ''It's a small change but irritating because I don't know why it happened'' (FI-31). These freelancers often experimented with different tactics that could change their score in hopes of increasing or maintaining their high score. Increasing their score was particularly critical if it hovered around 90 percent because, as this top-performing freelancer conveyed, ''once you go below 90 percent, you are screwed, no one will hire you,'' because one's visibility in clients' search results was affected noticeably (FI-35). Almost every client I interviewed said they looked only at freelancers who had a score above 90 percent.

Experimenting with contract length. One way that high performers experimented with influencing the algorithmic evaluation after experiencing a setback was by altering whether they worked on short- or long-term contracts. One freelancer described working on a long-term project with a client for many months. The freelancer and client both expressed satisfaction in the completed project, but the freelancer shared that the client ''didn't leave any feedback even after I asked them'' (FI-32). The freelancer believed the algorithm penalized their evaluation score as a result. This experience left a sour taste in the freelancer's mouth: ''I no longer do long-term projects. . . . I do short-term projects so that I can get a lot of ratings, in case one of them goes bad or the

[11] Table 2 likely underrepresents the extent to which each experience occurred, in part because I followed an ethnographic interview approach, which encourages open, spontaneous responses (Spradley, 1979), rather than collecting information about specific experiences and outcomes. Some freelancers mentioned these consequences during the semi-structured interviews, while others did not say anything about them, as I left it up to them to share their experiences as authentically and comfortably as they felt they could. Despite the fact that not all freelancers spoke overtly about the consequences for which I coded, Table 2 illustrates the prevalence of these experiences for freelancers in my sample without explicit prompting. An analysis of how these consequences varied by each interviewed freelancer is available by request.

Figure 3. Overview of High and Low Performers' Reactivity Pathways

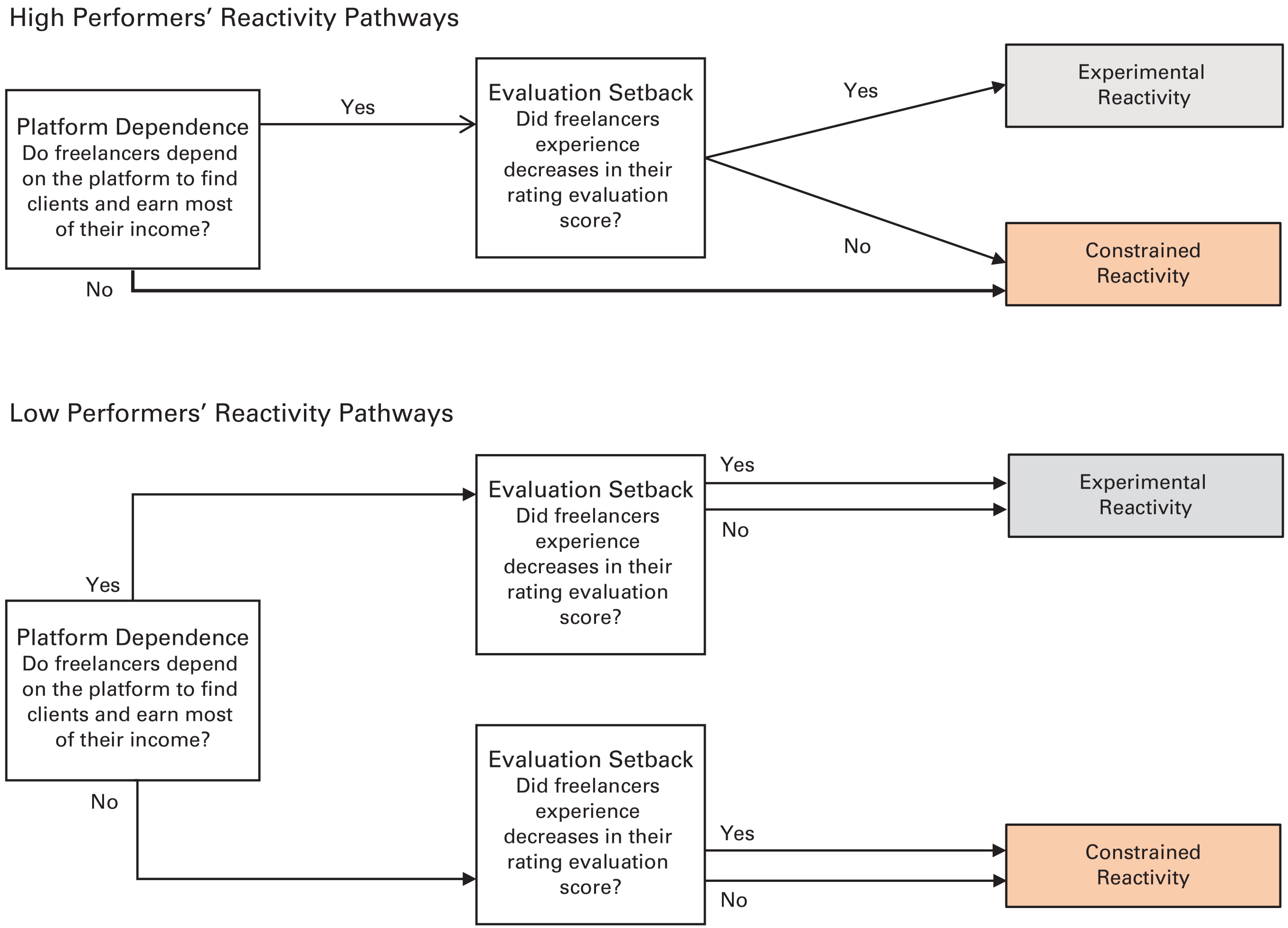


Table 2. Consequences and Reactivity of Opaque Third-Party Evaluation (Based on Interviews)

	Overall (n = 62)	Score ≥ 90 (n = 36)	Score < 90 (n = 26)	New Freelancers (n = 18)
Difficulty in learning from opaque evaluation	54 (87%)	31 (85%)	23 (88%)	*
Paranoia	36 (58%)	22 (61%)	14 (54%)	*
Experimental reactivity	46 (74%)	25 (69%)	21 (81%)	*
Constrained reactivity	16 (26%)	11 (31%)	5 (19%)	*

* Although TalentFinder collected data for its algorithm as soon as freelancers joined the platform, the platform did not provide new freelancers with an evaluation score until an unspecified number of jobs was completed. As a result, none of the freelancers I interviewed who were new to the platform were substantially aware of the evaluation algorithm.

client doesn't leave a rating'' (FI-16). Working on multiple short-term projects might buffer the impact of one project not receiving a feedback score.

Other high-performing freelancers' experiences led them to experiment with the opposite approach: working on long-term rather than short-term projects. This freelancer described being hired by a client who needed help building a research presentation for a rapidly approaching conference:

> Somebody asked me to prepare a [short] presentation. After I did it, they said they no longer wanted my services. I remember this clearly, because my evaluation score dropped. . . . Even though there are a lot of short-term contracts, I am not going to apply for them. I only sign up for long-term contracts. (FI-29)

The freelancer explained that this evaluation setback led them to believe that clients who had short-term contracts were risky because they could take a freelancer's work and end the contract without explanation or compensation. While this could theoretically happen on any project, the freelancer explained that a short-term project did not offer enough time to establish a relationship with the client to prevent such an outcome. To them, a long-term contract signaled that the client was willing to establish a relationship with the freelancer and was more invested in the project.

Experimenting with contacting clients to close projects. High-performing freelancers also experimented with whether they should approach a client toward the end of a project and explicitly ask them to close a contract and leave feedback. After experiencing an evaluation setback, a freelancer recalled that ''yesterday I had to message two clients explaining why I was asking them to leave feedback for the third, fourth, fifth time. It's so embarrassing and makes me look like a whiny, unprofessional freelancer who only cares about her stats!'' (FI-25). Even contracts that did not move forward and involved no exchange of work or money received this kind of attention. One example of this occurred when a freelancer explained that a client initially invited them for a project but subsequently decided they no longer wanted to work with the freelancer. A contract had been initiated, but no work had been completed, and the client had not paid the freelancer. Because the freelancer believed that a similar experience had resulted in an evaluation setback, the freelancer sent the following message to the client: ''TalentFinder is mysterious and does not show the algorithm's calculations, but can we log one hour and end the contract after that?'' (FI-41).

Other high-performing freelancers dependent on the platform experimented with closing contracts on their own rather than contacting clients. A freelancer on the discussion board recounted:

> I had almost 11 open jobs and new clients got hesitant to hire me thinking I am too busy, whereas five of those contracts were complete and despite various reminders, the clients did not close the contracts or [leave] any feedback. Therefore, I ended three contracts myself. . . . I did not receive any feedback on the [closed] contracts . . . but to an extreme shock to me, my feedback score declined.

Experimenting with platform activity levels. High-performing freelancers who experienced evaluation setbacks also experimented with actions related to their level of platform activity in an attempt to maintain or increase their scores. They did so out of concern that TalentFinder might punish their decreased presence on the platform. A freelancer who had used the platform since its inception offered an example of this kind of concern. They had secured a full-time, three-month project with a local employer (not on TalentFinder), and after completing it, they logged back on to TalentFinder. This freelancer's score had ''decreased by over 20 percent'' in those three months (FI-46), and TalentFinder had marked their profile as ''private'' so that it no longer appeared in clients' search results. When this happened to a different freelancer, they believed the algorithm penalized their rating score every week because ''I did not log onto my profile, [and] all of a sudden I am invisible'' (FI-53) to clients. In response, this freelancer began logging on to TalentFinder every day in the hope that increased activity might preserve their score. Another freelancer I interviewed took it a step further: even though they had secured a full-time job, they still took ''small jobs'' on the platform regularly just in case their evaluation score decreased due to inactivity—and in case they needed to return to TalentFinder for project work (FI-15).

These freelancers experimented because they could not be certain which data TalentFinder's algorithm was collecting and how it was being used. Their concerns seemed warranted because TalentFinder specifically mentioned in the terms of service:

> We and our third party service providers, including analytics and third party content providers, may automatically collect certain information from you whenever you access or interact with the [platform]. This information may include, among other information, the browser and operating system you are using, the URL or advertisement that referred you to the Service, the search terms you entered into a search engine that led you to the Service, areas within the Service that you visited, which links you clicked on, which pages or content you viewed and for how long, other similar information and statistics about your interactions, such as content response times, download errors and length of visits to certain pages and other information commonly shared when browsers communicate with websites. We may combine this automatically collected log information with other information we collect about you. We do this to improve services we offer you, and to improve marketing, analytics, and site functionality.

Each type of experiment freelancers tried had uncertain outcomes. As explained by a high-performing freelancer who had completed over 160 jobs on the platform and read every article, blog, and discussion board post on the algorithmic evaluation they could find, ''The exact formula behind the calculation of the [score] seems to be one of the great mysteries of our time. We don't really know how all the factors are weighed, and there even seems to be a dispute about which numbers factor into the rating algorithm at all'' (FI-6).

Constrained Reactivity: When High Performers with High Platform Dependence Do Not Have Evaluation Setbacks

When high-performing freelancers who depended on TalentFinder did not experience an evaluation setback, they responded to the difficulty learning from the evaluation and the paranoia they experienced by limiting their exposure to it. They disengaged from the platform in hopes their actions decreased their chances of experiencing an evaluation setback. As one freelancer speculated, ''if the algorithm has nothing to record, it can't ding my score'' (FI-12). Their constrained reactivity consisted of working with their TalentFinder clients off-platform and refraining from working with new clients.

Working with clients off-platform. Rather than experimenting with new tactics, some freelancers responded to the opacity of the algorithmic evaluation by working with TalentFinder clients outside of the platform. When clients and freelancers met through TalentFinder, the platform barred them from taking their payments off the platform for at least two years. Per TalentFinder's user agreement, clients and freelancers agreed ''to make and receive payments only through TalentFinder for two years from the date you first meet your Client or Freelancer on the Site.'' The user agreement specified several penalties associated with violating this term, including the possibility of account deactivation. The agreement also required users ''to notify TalentFinder immediately if a person suggests to you making or receiving payments outside of the Site.''

Because of these terms, and to ensure I did not expose freelancers or clients to undue risk and potential suspension, I did not ask informants explicitly whether they took work off-platform. Nonetheless, many high-performing freelancers who did not experience an evaluation setback but depended on the platform to find work conveyed that they requested that clients arrange projects and payments outside the platform. This occurred when a high-performing freelancer, who I invited for an interview, found my private e-mail address online and wrote:

> We can talk—we just can't do it via TalentFinder, if that makes any sense to you. They don't allow me to ''work for free'', and technically I am not even allowed to talk to you outside of the platform since we connected there. But since I don't want you to hire or pay me, I don't care.
>
> I'll gladly ''donate'' an hour of my private time. (FI-14)

In our conversation, the freelancer explained that they wanted to speak with me, but they were unsure how the algorithm would interpret our interaction and how it would affect their evaluation score, since the monetary value of the contract was less than their typical project. As such, even though they expressed confidence that I would give perfect evaluation scores, and they had not experienced an evaluation setback, they felt it was safer to conduct the interview off-platform. This person refused to take any payment to ensure the interview could not be construed as taking work off the platform. Seven clients I interviewed also voluntarily shared that freelancers had requested working with them outside TalentFinder. Because clients, as one high-performing freelancer described, "do not know anything about this [algorithmic evaluation]" (FI-4), freelancers pitched the prospect of working off-platform as a win–win, since each party could save the fees that TalentFinder normally extracted.

Freelancers explained that working outside of TalentFinder allowed them to focus on completing their work: they did not have to worry about the quirks associated with the algorithm, especially if a project did not work out. As one freelancer wrote on the discussion board:

> I've also worked as an independent contractor outside of TalentFinder for 17 years, and I have an excellent reputation. Have I had a couple of relationships that didn't work out? Sure. Do I speak badly about them? No. Do they speak badly about me? No. Do I refer potential new clients to them for a good review? Hell no! But I also don't wear a scarlet "F" around my neck as a result of these jobs that didn't work out.

Reluctance to work with newcomers. Because many new clients lacked experience working with freelancers in a platform setting and often had naïve expectations about what freelancers could accomplish, high-performing freelancers who depended on the platform but did not experience an evaluation setback shied away from working with them. As a freelancer expressed on the discussion board, a "novice client with outrageous expectations due to lack of experience can ruin a freelancer's reputation on a whim with no human oversight from TalentFinder's flawed algorithm." Another freelancer on the discussion board shared, "We're punished for what new clients don't understand. They have no idea about leaving ratings and feedback, don't understand payment systems well, don't understand the platform well as it's not intuitive." As a TalentFinder client, I experienced firsthand freelancers' reluctance to engage with new clients. A high-performing freelancer, who had not experienced an evaluation setback, waited close to two months after I invited them for an interview because they "had to make sure you were reliable by seeing how you [publicly] rated other freelancers" (FI-22). Another freelancer added, "I don't waste my time with new clients . . . they have no idea how to use TalentFinder."

Constrained Reactivity: High Performers with Low Platform Dependence

High-performing freelancers who had low dependence on the platform engaged in constrained reactivity regardless of whether they experienced evaluation setbacks. One such freelancer told me, "I don't have time to keep up with all of the changes that happen with the rating algorithm . . . it's too much to keep track of" (FI-19). Because these freelancers had active projects outside of TalentFinder and, at the time, had what TalentFinder considered a high score, they did not feel compelled to experiment on the platform. Instead, they felt more comfortable sticking to their current approach when working on the platform. For instance, a freelancer on the discussion board expressed:

> I am not going to approach them [i.e., clients] and say "Can you hurry up on this project [and leave a rating] so TalentFinder won't whack my 'Success Score'?" That

> would feel wholly unprofessional to me and I won't do it. I am working hard with each and every project to do my very best and I am having success.

A high-performing freelancer who had low dependence on the platform and experienced an evaluation setback shared, "I was shocked to see my score decrease," but rather than experimenting with new approaches, "I just went back to my [off-line] project" (FI-23). Others were content to wait for the right client and project to come to them, as this freelancer expressed: "ever since I became top-rated, I don't have to apply for projects anymore . . . clients reach out to me all the time" (FI-28). Even though this freelancer considered the algorithmic evaluation to be like "black magic" because of its opacity, they were prepared to wait until a client on TalentFinder contacted them, and they chose not to worry greatly about possible consequences for their score.

Low Performers' Reactivity Pathways

I found that the reactivity pathways of low performers—those with scores below 90 percent—were primarily influenced by whether they were dependent on the platform; evaluation setbacks did not have a noticeable role in how they reacted. If they had high platform dependence, they experimented with ways to increase their scores regardless of whether they experienced an increase or decrease. But those who did not depend on the platform gradually withdrew from it because of the difficulty of determining how the algorithmic evaluation might change their scores. The findings below describe low-performing freelancers' experiences and their reasons for engaging in experimental or constrained reactivity.

Experimental reactivity: Low performers with high platform dependence. A freelancer who depended on the platform whose score was below 90 percent shared that "all I can think about is figuring out how to raise my score" (FI-47). They explained that without a score above 90 percent, "it's as if no client wants to look at you." But increasing their score was not straightforward given the algorithm's opacity. Another freelancer said, "I recently finished a long-term job with a client . . . everything went perfectly, but my score didn't budge. . . . I have no idea how the thing increases" (FI-45). These freelancers experimented to try to increase their scores, and their experimentation at times went further than that of high performers, in hopes they could increase their scores more quickly. A freelancer who normally applied for software engineering jobs shared, "After my rating decreased, I started to apply for admin jobs" because "they are faster, and there is less risk that something could go wrong when the client will give you a bad rating" (FI-52). This freelancer thus not only experimented with working on projects with different contract lengths but with taking jobs in a different job category and skillset.

Freelancers who experienced score increases but remained below 90 percent continued to experiment because "it takes a moment for it to drop but takes forever to increase it" (FI-39). They had to try harder to convince clients to hire them instead of top-scoring freelancers, and they sometimes experimented with contacting clients: "I do a lot of research about the client and customize every message to them, instead of sending out a generic 'I am interested in your project' message. . . . some clients really appreciate that personal touch, and it helps me stand out" (FI-48). Freelancers did not enjoy having to experiment, but as this freelancer conveyed, "I had no choice but to keep trying [to increase their score]. . . . I needed work" (FI-52).

Constrained reactivity: Low performers with low platform dependence. Low-performing freelancers who did not depend on the platform became disengaged from it instead of trying to figure out how to increase their scores. As one freelancer shared, "I stopped logging on once I saw my score decreased so much" (FI-60). Another who had tried out TalentFinder as a bridge to finding an offline job told me that "people don't hire you when it [i.e., evaluation score] is that low," and because "figuring it out [i.e., increasing their evaluation score] is a pain in the butt, I stopped applying for jobs" (FI-50). People in this group who did experience increases in their scores were not persuaded that such increases were sustainable, as this freelancer shared: "it takes me minimum three months, and it goes up 5 percent maybe, if that" (FI-51). Another worker shared that "the negative feedback takes you back for months," so rather than experimenting with different ways to gain projects and increase their score, "it's not worth all the hassle. . . . I moved on to find projects in my town" (FI-43).

Table 3 shows representative quotes by year from the discussion board data, illustrating the sustained impact on workers of TalentFinder's choice to implement the opaque algorithm. While the new evaluation system was

Table 3. Representative Quotes of Consequences and Reactivity from Discussion Board Data through Time

6 Months or Earlier	1.5 Years after Introduction	2.5 Years after Introduction
	Difficulty in learning from opaque evaluations	
"I had 92% score even last week. By this time I haven't closed any job. But this week the algorithmic evaluation automatically downgraded to 90%! 2% minus without any reason! Can anybody tell me why it happened? What I should do now? Thanks." "My score is at 68% but I see that there is only 1 negative comment in an otherwise positive feedback from a client; all my other feedback is positive. What can I do?" "Today my Score has been updated. Before the update, my Score was 61%. In the last 15 days, 3 job have been ended. The clients who ended contract, gave me very good feedback, but I am wondering why my Score remains the same that means 61%. I cannot find any cause for this [lack of change in the score]."	"The main problem I see with the algorithmic evaluation score is that it doesn't help you in any way to improve. You don't know what you are doing wrong and for which project. if at least it gave you some hints like, 'you should try to deliver something every week,' 'don't left jobs without activity for more than 3 days.'" "While I'm ok with my score, I have no way of knowing if it will drop or increase since the calculations behind it are secret. I find that very strange. How does that protect the 'integrity' of the algorithmic evaluation?" "I am facing issue with my score, which was 100% from the beginning. But recently it dropped down to 95% and surprisingly I never got less than 5 stars or a bad review in any of my contracts. You can visit my profile and see all of my client's feedback on my work and I never removed any	"Hello everyone, This is [redacted] and I have been on TalentFinder for long time period. I have had a 100% score but it has been down to 71% although there is no specific reason. (Can anyone) please share your findings on why it could be? Thanks in advance." "My Score, before yesterday, was 97% and I was quite happy about it. However, all of sudden, I found out today that my score has fallen to 91%. It is especially confusing because none of my contracts ended in the last 15 days. Can someone please guide me what has caused this massive fall in my evaluation score?" "I want to know what happen to my profile? I had 92% score, my success rate goes to 87% suddenly? what did I do wrong? I collected 5 stars ratings in last 2 weeks and I don't have any lower than the one with 4.7, the rest with 5 stars. This is not fair.

(continued)

Table 3. (continued)

6 Months or Earlier	1.5 Years after Introduction	2.5 Years after Introduction
	feedback ever. I am so disappointed with it as I completed all contracts with good feedbacks and all my clients ended the contract on a good note and some of them contacted me again for more work. Can anyone tell me why my success score dropped? And is there any way to get my success score back to 100%?''	Please, could anyone help with explaining what I'm doing wrong?''
	Paranoia	
''Allow me to get paranoid for a second: I work at translating documents. If I do a search and I find 2 other translators with a 100% score and I want to hurt their score, TalentFinder allows me to hire (or have a buddy of mine do it), them and give them an 'I will not hire again' private rating at the end. This will lower their score and allow [me] to look like a better option to the client. The same system that is supposed to protect me is also allowed to hurt me.'' ''TalentFinder is fiddling with something. The majority of my clients are long term. Am I being penalized for few recent hirings? Computers have done stranger things.'' ''I really need to know what's going on here [with the algorithmic evaluation], because now I think I have reason to think you all [TalentFinder] are scamming people.''	''Have others noticed a significant slowdown in successful bids/ interviews coming in lately? I've been applying like crazy and since my score is 95% and I have some really great client reviews, those things combined usually result in at least a few new jobs coming in. Unfortunately, for me it's pretty much nothing but dead silence. . . Maybe I'm just being paranoid and hopefully, the tides will turn back my direction soon.'' ''I have read all information available in customer support about the algorithmic evaluation, but I still don't know how just one bad client in one project can take 11% from my score. Looks like very weird maths are working on those algorithms.'' ''Don't you see the absurdity in all of this? I shouldn't be doing months of extra work just to undo some unpolished algorithms. So you're basically saying there's nothing I can do, even though it's arguably none of my fault, and my score plummeted.''	''Instead of focusing on delivering great work to my clients, I will now have to focus more on dodging curveballs that TalentFinder is apparently throwing at me.'' ''Waking up to a 70% job success rate when you have a history of purely happy clients (happy to your face at least) is not only unfair, it's really just a slap in the face for doing really good work. It is an arbitrary grading system and is almost heartbreaking to think that the reputation you spent months or years building up is suddenly and inexplicably ruined.'' ''Many people have found significant drops in their score today. Mine only went down one point, but it's strange because I completed a couple jobs with great feedback, so I'm watching it closely. There was some thought that there may be a bug since some people saw 2 drops in their score in a single day, when normally the calculation is refreshed only once in a cycle.''
	Experimental reactivity: Experimenting with contacting clients to close contracts	
''I'd try and get as many of them (i.e., contracts) closed by the clients as possible now, and then close the ones left one at a time with contracts that closed with feedback in between.'' ''Personally, I think it is better to leave a job open for a while and hope the client will come back and close it with feedback.'' ''Try and get clients to end contracts more or less as soon as the work was done—the longer you leave it the less invested is the client.''	''When I finish a fixed-price task, I DO NOT send a message to the client. I immediately click the 'Send Work and Request Payment' button. That is all.'' ''I did the work and the client paid it in full (fixed price project). He liked my work, but he forgot about the feedback and closing the contract and left. Then I myself closed the contract.'' ''I've just noticed that my evaluation score has been reduced . . . after a client closed the contract. . . . My score was affected immediately after that.''	''I just observed that the 'zero contract' has not yet any negative impact on my evaluation score.'' ''I've gotten to the point where I'm about to cancel some hourly contracts due to the client disappearing and not responding to messages (for well over a couple months).'' ''I closed one old contract with an unresponsive client and my evaluation score dropped from 100% to 88%.''

(continued)

Table 3. (continued)

6 Months or Earlier	1.5 Years after Introduction	2.5 Years after Introduction
	Constrained reactivity: Taking work off of platform	
''I am so fed up with this system that I've shifted some stuff over to another website, some clients are now private and I'm trying my best to move a load of other clients.'' ''3 months ago I was a 'Top Provider' and I believe this was done to most great providers who have a long history of work here . . . then they yanked it away by implementing a very damaging algorithm. If this is any indication of where TalentFinder is heading, I prefer to go where real work is done by real providers.'' ''I am fed up and will be steering clear from TalentFinder. Be careful for all freelancers, they are all just playing you and milking you, you can fall ANYTIME with NO REASONS AT ALL.''	''I do not like to complain about the algorithmic evaluation but some things are still odd and need a further explanation beyond 'we hide the algorithm to prevent cheating.' For the first time . . . I have an impression that TalentFinder wants to force a large number of freelancers out.'' ''Someone said not to 'obsess over your score.' That's very hard to do. I noticed my score plummeted from 100% to 87% after I had not received responses on proposals for a while . . . Why am I being punished so harshly for allowing clients to take their time compiling content? Please advise if this is currently reversible if I close all projects asap? or is it too late? I want to continue using TalentFinder, but such an injustice is making it really difficult and demoralizing.'' ''I cannot figure out the why on earth with no outwardly public criticism or even a negative comment during communications, that I have a 70%, as this is fundamental in the way TalentFinder functions. I have decided to stop. Yes, I know me leaving TalentFinder won't make a sweet bit of difference, but if everyone that is pissed off with the algorithmic evaluation does as I am doing, someone, somewhere at some point will think, 'hold on a minute.'''	''I've done lot of sacrifice to keep my score intact but unfortunately it dropped today for no apparent reason especially when all my clients are happy. I've done nearly one hour chat with one of the [TalentFinder] agent and she/he didn't come up with a specific reason why it drops. This really frustrating no one can work in an such environment, and I'm sure lot of great freelancers leaves TalentFinder cause of this. For me, I'll finish my current contracts (in favor) of my great clients and I'll just leave.'' ''As a 3-year freelancer with 300 jobs under my belt you are doing everything you can to encourage me to leave the site. I've dramatically reduced my jobs on TalentFinder as the feedback system favors the client rather than looking after the freelancer.'' ''There's loads of freelancers who are pissed off with the algorithmic evaluation, yet the answer is always the freelancer. Enough is enough. I am done with TalentFinder.''

successful in limiting the number of freelancers who fell in the top performance category to 5 percent of the total freelancer population on the platform, which made it easier for clients to differentiate workers based on their rating scores, I saw evidence that TalentFinder was concerned about the constrained reactivity practices freelancers enacted. During my study period, TalentFinder implemented an algorithm to detect when freelancers and clients exchanged contact-related or other information through their messaging service that could lead them to work off-platform. For example, after I sent a freelancer a meeting link to interview them outside the platform, I received this automated message: ''Communicating outside of TalentFinder before a contract has started or paying outside of TalentFinder is against our Terms of Service and could result in account suspension.'' Such efforts suggest that TalentFinder was aware of and concerned about reactivity practices leading to less engagement on the platform.

DISCUSSION

Contributions to Our Understanding of Control

People experience third-party evaluations as a form of control because these evaluations are consequential for their access to resources, recognition, and opportunities provided by external audiences (Orlikowski and Scott, 2014; Espeland and Sauder, 2016; Brandtner, 2017; Kornberger, Pflueger, and Mouritsen, 2017). In the TalentFinder context, workers experienced third-party evaluations as a form of control because they determined their visibility and success on the platform, but the platform obscured both the criteria influencing the algorithmic evaluation and the rate at which these criteria changed. Because freelancers, based on their experiences with more traditional evaluations, expected to use their rating scores to learn how they could better align their actions with the evaluation criteria, they found their inability to learn from the opaque system deeply frustrating. Their ongoing inability to learn and frustration from the opaque evaluation, even years after its introduction, led workers to feel paranoid because they could not determine what actions would influence their rating scores.

An apt metaphor for workers' experience with opaque algorithmic evaluations is an invisible cage. This metaphor highlights that workers experience such evaluations as a form of control but cannot clearly see or understand how they work. While previous research has highlighted the uncertainty workers encounter with opaque third-party evaluations more broadly (Orlikowski and Scott, 2014; Shapiro, 2017; Rosenblat, 2018), this study delineates components of the unpredictability workers experience with opaque evaluations. These components include *evaluation criteria*, because they can shift and are largely opaque; *execution of the evaluation*, because the intervals at which evaluation changes occur are uncertain; *magnitude of the evaluation*, because the extent to which the evaluation changes cannot be identified; *impact of the evaluation*, because workers are unsure how changes in their scores affect their short- and long-term reputation and ability to secure further work; and *who influences the evaluation*, because the extent to which each party (i.e., client vs. platform) affects the evaluation score is unspecified. Delineating these components illuminates why freelancers have difficulty aligning their behavior with the evaluation's criteria, even in the long term. That is, even if a freelancer makes headway in uncovering one or a few components of the evaluation algorithm, other facets will remain opaque. As a result, workers inside the invisible cage have difficulty accessing platforms' expectations for success and find these opaque and changing without notice, explanation, or recourse.

What are the broader implications of the invisible cage? Facilitated by the explosion of data and algorithms, labor platforms are attempting to "see" people in a new way (Fourcade and Healy, 2016). Markets and organizations have always tried to classify people into broad, aggregate categories. Previously, such classifications were based on characteristics such as income, gender, age, and race. Now, "the new classifier is inside, looking around. It knows a lot about what you have done in the past. Increasingly, the market sees you from within" (Fourcade and Healy, 2016: 23). Seeing "from within" means that platforms employ algorithms to collect users' data as they interact on the platforms, dynamically classifying people using various ratings, rankings, and categories; the users cannot verify what data were collected and may struggle to understand how or why they were categorized a given way (Finn, 2017; Zuboff, 2019). Because only a platform's algorithms can crunch the relevant data and make sense of how users' actions translate into evaluation outcomes, algorithms purport to "know" the individuals in the invisible cage in ways that are inaccessible to them.

This phenomenon has become increasingly common: from the credit card industry to health insurers and retail stores, many platforms, organizations, and markets use algorithms to collect people's data and see people from within, with minimal external scrutiny or recourse for those being evaluated (Pasquale, 2015; Fourcade and Healy, 2016; Finn, 2017; Zuboff, 2019; Heaven, 2020). For example, a recent report identified more than 120 companies, ranging from technology and advertising to cybersecurity and tenant-screening companies, that buy and sell people's data collected by third-party platforms and use it to categorize and make consequential decisions about people in ways that are typically inaccessible to those whose data was collected (Pasternack and Melendez, 2019).

Much like the iron cage metaphor dominated our understanding of workers' experiences in bureaucratic work settings (DiMaggio and Powell, 1983), this study's findings suggest that the invisible cage metaphor is apt for characterizing workers' experiences in labor platforms specifically and the gig economy broadly. Increasingly, workers encounter opaque algorithmic systems that change at a rate that affords the people affected minimal insight or recourse.

Contributions to Our Understanding of Reactivity to Third-Party Evaluations

This study sheds light on workers' reactivity to opaque third-party evaluations and on mechanisms that contribute to their reactivity in platform settings. Prior research has suggested that regardless of their performance or ranking, people will try to engage in convergent reactivity such that their actions align with the third-party evaluation's transparent criteria and thus increase their chances of receiving favorable outcomes from external audiences (Sharkey and Bromley, 2015; Espeland and Sauder, 2016; Brandtner, 2017). Labor platforms, however, have increasingly made their evaluations opaque to workers to more easily differentiate among their evaluation scores when presenting them in search results to potential clients (Garg and Johari, 2020). This study shows that when they are subjected to opaque algorithmic evaluations, workers have difficulty identifying the evaluations' underlying expectations, including what is rewarded or punished; to what degree various actions affect them; when the effects will transpire; and exactly who or what caused the effects (Briscoe and Murphy, 2012). While prior literature examining opacity at the organizational level (Proctor, 2008; Briscoe and Murphy, 2012) would suggest that freelancers would struggle to understand how to react to such opacity, it does not provide insight into the situated responses and mechanisms contributing to freelancers' reactivity. This study demonstrates how the difficulties freelancers experience with opaque third-party evaluations contribute to variation in the types of reactivity they enact, and it reveals mechanisms leading to variation within and between high and low performers' reactivity.

My findings show that platform dependence and evaluation setbacks are key mechanisms contributing to variation in workers' reactivity. Results highlight that the importance of third-party evaluations to workers on platforms varies depending on their platform dependence, and this variation can determine their responses to such evaluations. For instance, high-performing workers who benefit from increased visibility and access to opportunities on a platform may nonetheless constrain their involvement with it if they do not depend on the platform to find work and wages (Schor et al., 2020).[12] Someone in this position may enact constrained reactivity because deciphering the evaluation criteria and keeping up with its changes may seem too cumbersome to be worthwhile.

For high performers with high dependence on a platform, in contrast, whether they experience evaluation setbacks (score reductions) is key to variation in their reactivity to opaque third-party evaluations. Those who experience evaluation setbacks are likely to enact experimental reactivity, while those who experience no setback will tend to enact constrained reactivity. The study results suggest this difference occurs because when freelancers experience evaluation setbacks, their visibility in clients' search results can diminish, increasing their difficulty with finding work on platforms. High performers have more to lose than low performers when experiencing such setbacks. Thus high performers who experience an evaluation setback are likely to increase their platform activity and experiment with ways to increase their rating scores to retain the benefits that come with high-performing status. Those who do not experience an evaluation setback are not likely to take such actions given that any new actions could introduce the risk of decreasing their scores. Thus those who have not experienced an evaluation setback are likely to enact constrained reactivity practices in hopes that limiting their platform activity and exposure to the algorithm may protect them from rating setbacks. Further, while I found that both high and low performers enacted experimental reactivity, low performers are likely to engage in a more wide-ranging set of practices, such as taking projects outside of their main job category, because they hope to more quickly increase their scores.

There is emerging evidence that the two mechanisms identified in this study, platform dependence and evaluation setbacks, lead to experimental and constrained reactivity in other platform settings. Caplan and Gillespie (2020: 7), for example, noted that content creators described YouTube as making "capricious, ad hoc, and unpredictable changes in the rules, [and] also about YouTube failing (or avoiding) to communicate the rules and expectations clearly." Such changes have motivated content creators who are dependent on the platform for their wages and have experienced setbacks in their visibility on the platform to experiment with tactics to increase that visibility, including attempts to artificially inflate viewership counts. Caplan and Gillespie (2020: 8) even noted that a content creator "reported running an experiment" on the platform to see how different keywords affected the algorithm controlling the visibility of one of their YouTube videos. In another example, Amazon delivery drivers experimented with ways to induce the work-assignment algorithm to prioritize assigning work to them, including hanging cell phones on trees near

[12] Even in non-platform settings, when actors are subject to relatively transparent third-party evaluation systems, we see evidence that if dependence on the evaluation is low, high performers may enact constrained reactivity. See, for example, Harvard, Stanford, and Wharton opting out of recent MBA ranking lists: https://www.businessbecause.com/news/mba-rankings/7468/financial-times-mba-ranking, last accessed Feb. 9, 2021.

the package distribution center to appear ''closer'' to it (Soper, 2020). Resonating with the constrained reactivity observed in my study, Curchod et al. (2019: 662–663) found that high-performing eBay sellers who depended on the platform for revenue but wished to avoid its algorithmic evaluation found ways to interact with buyers off-platform. My findings extend these emerging insights to identify and theorize how these mechanisms contribute to variation in both high and low performers' platform-related reactivity more broadly.

Existing studies have highlighted how mechanisms such as legitimacy, commensuration, and self-fulfilling prophecies contribute to actors' convergent reactivity when subject to transparent third-party evaluations (Sharkey and Bromley, 2015; Espeland and Sauder, 2016; Brandtner, 2017). My research extends our understanding to platform settings, demonstrating that platform dependence and evaluation setbacks are key mechanisms contributing to variation in reactivity when actors are subject to opaque third-party evaluations.

Contributions to Our Understanding of Evaluations

Recent studies have highlighted the importance of developing theories reflecting the conditions workers encounter in platform settings, which have become a defining feature of the gig economy (Vallas and Schor, 2020; Lei, 2021; Stark and Pais, 2021). Curchod et al. (2019) showed how eBay uses evaluations to unilaterally impose its expectations on users, showing that evaluations serve as disciplinary tool for platforms as they did for traditional organizations—but to an even greater degree. My findings showcase a different role of evaluations.

In traditional bureaucratic settings, evaluations have always played an important role in ''tightening the iron cage,'' because they allow management to explicate expectations and assess alignment of workers' behavior with these standards (Edwards, 1978; Curchod et al., 2019). Managers responsible for filling out workers' evaluations in such settings are aware of the criteria in play and how their decisions will affect workers' success (Castilla and Ranganathan, 2020). In the traditional context, informal, opaque inputs such as biases and discrimination exist, but they are viewed as problematic (Castilla, 2008; Castilla and Benard, 2010; Rivera and Tilcsik, 2019) and are subject to established internal (e.g., human resources) and external (e.g., litigation) pathways for recourse. Many U.S. states, for example, by law allow employees to view their employment records, which include their written evaluations.

The platform I studied did not provide freelancers with a pathway to appeal or modify their scores. Further, clients were largely unaware of how their ratings affected freelancers' platform visibility and scores. In fact, many clients believed their private ratings did not contribute to freelancers' scores—that only their public ratings did. This belief contradicted freelancers' experiences and available evidence. The comparatively limited knowledge clients had about their role in influencing freelancers' platform success stands in stark contrast to the central role managers knowingly have in evaluating workers in traditional organizations.

One reason clients were relatively uninformed was that they had a largely transactional relationship with the platform and freelancers. Clients, unlike freelancers, did not need to understand how the platform operated and did not have to interact with freelancers after they considered a project complete and provided their public and private ratings. Moreover, clients faced few or no repercussions if they did not pay a freelancer, close a contract, leave a rating, or perform virtually any other action on the platform. Thus my findings highlight that in a platform setting, the purpose of evaluations has shifted from ensuring workers' behavior is aligned with organizational expectations to optimizing platforms' outcomes, such as through providing information to clients, with little to no opportunity for workers to learn and improve based on their evaluations.

Implications of the Rise of Algorithmic Evaluations

This study also highlights at least two fundamental differences between algorithmic evaluations and traditional evaluations. First, as my data and those of other studies have shown, algorithms can evaluate and adapt at an unprecedented scale and speed compared with traditional evaluations, which helps maintain opacity and avoid scrutiny (Orlikowski and Scott, 2014; Dourish, 2016). Even if people reverse-engineer or understand an algorithm at a moment in time, algorithms can update instantaneously based on real-time data, effectively rendering the current understanding of the algorithm obsolete. This reality makes auditing and regulating an algorithm particularly difficult, even if an individual or agency has access to its underlying procedures (Pasquale, 2015; Dourish, 2016; Finn, 2017). As many scholars have noted, for example, despite the intense scrutiny over the influence of platform-based algorithms on past elections and other high-stakes events, we have little knowledge about how those algorithms operated (Gillespie, 2018).

Second, and relatedly, this study highlights that algorithms lack reflexivity despite the speed and scale at which they operate; they simply carry out their programmed instructions regardless of the situation (Alkhatib and Bernstein, 2019). An algorithm's lack of reflexivity becomes apparent when it encounters novel situations that are not part of or represent only a small portion of its training data. In such situations, the algorithm will still apply its encoded procedures, sometimes with unintended social consequences (Lambrecht and Tucker, 2019). In this study, for instance, I observed that dynamic when a client inadvertently provided an erroneous score for a freelancer. The algorithmic evaluation appeared to interpret the entry as a valid, negative signal and carried out its programmed instructions, which had immediate repercussions for the freelancer's visibility on the platform. In a similar circumstance, a human providing an evaluation could easily recognize a mistake in the moment and adjust their actions, such as modifying a feedback score, without receiving additional training and data. An algorithm's lack of reflexivity underscores the situations in which algorithms excel (i.e., familiar, predictable situations) and highlights why they have struggled when applied to social settings, which are inherently unpredictable and complex (Alkhatib, 2021).

Boundary Conditions, Limitations, and Future Work

This study reveals important boundary conditions and limitations, which are important to consider for future studies examining the impact of opaque third-party evaluations. First, TalentFinder operates a global, high-skilled platform with millions of users. In high-skilled work it is difficult to specify how people should behave, because the work is inherently ambiguous; it is challenging to predict how the work will unfold, how long it should take, or how to evaluate its quality (e.g., Rahman and Barley, 2016). Even well-planned projects may unfold unpredictably (Vaughan, 1996). My findings suggest that when a platform cannot specify how freelancers should behave—due to the nature of the work, to prevent gaming behavior, or for other reasons—the platform may be more likely to employ opaque third-party evaluations. Other platforms with more predictable, defined tasks seem likely to be more transparent with their third-party evaluations because the expected behavior is comparatively more straightforward.

Second, TalentFinder was considered the best labor platform for high-skilled work; it had no clear competitors. Freelancers noted few credible online alternatives offering quality clients or the ability to process payments to globally distributed workers. TalentFinder thus had no trouble attracting workers and even had to limit the number of freelancers who joined. In fact, a discussion board message posted in 2020 read: "TalentFinder is absolutely flooded with freelancers. Drowning in them, actually. There are way more than could ever get hired, and of the ones who are accepted, only a really small percentage ever wins a job." Lack of competition may influence which platforms employ opaque evaluations. We have numerous examples of the largest platform or entity in a market space enacting changes that are in its own interests but harmful to end users (Chen, 2016; Allcott and Gentzkow, 2017; Morris, 2017; Kailath, 2018; Keller, 2018). Third, TalentFinder and other platforms are clearly operating in an emerging space in which governments and legal systems are struggling even to define the employment relationships between platforms and their users, let alone regulate them (Khan, 2017). Freelancers thus currently have few options for legally challenging the conditions they encounter.

Finally, this work presents a qualitative study of a single large labor platform. While the goal was to relate a wide range of freelancers' experiences, each of my data sources involved unavoidable selection biases. Though I tried to attract a wide range of freelancers and did not observe any evidence suggesting that certain types of freelancers were more interested in being interviewed, I cannot definitively say why certain freelancers were interested in participating. It is possible, for example, that the wage offered was more attractive to certain freelancers or that TalentFinder's algorithms displayed the job posting to specific groups versus others. The goal of collecting data from each source was to understand how freelancers made sense of their experiences on the platform, not to achieve statistically proportionate representation of the freelancer population to test a specific hypothesis, which mitigates these concerns. Further, although the data I collected were not necessarily representative of all freelancers' experiences and thus may contain bias, they provide insight into how freelancers reacted to the evaluation score regardless of an individual freelancer's potential bias, opinion, or experience with the algorithm. That is, all freelancers had access to the same information and were equally subject to TalentFinder's opaque evaluation process. Even if a freelancer was not bothered by the opaque nature of the algorithm, they could not access information that would help them learn how their or others' actions contributed to changes in their scores.

Even so, more work is needed to determine how the mechanisms described in this study generalize and how other important factors such as age, gender, and socioeconomic status affect freelancers' experiences and reactivity. Future studies can investigate such factors to help build a more complete picture of the impact of opaque third-party evaluations on freelancers' reactivity. And while this study highlights the prominence of opaque third-party evaluations on one labor platform, future work can assess their impact in other settings. Reports from ''surveillance states,'' for example, suggest that governments can enact both transparent and opaque evaluation systems at different times depending on their goals (Liang et al., 2018).

The prominence of opaque third-party evaluations is increasing, especially with the mounting use of algorithms and artificial intelligence (Finn, 2017; Liang et al., 2018; Zuboff, 2019). Yet the long-term impacts of such evaluations are unclear, especially in the context of shifting political and legal landscapes (Khan, 2017). This study contributes to a growing stream of research shedding light on workers' experiences in emerging settings. How gig workers react to opaque third-party evaluations seems increasingly likely to determine their success or failure on labor platforms and their likelihood of participating in such settings going forward.

Acknowledgments

I gratefully acknowledge the support and guidance by Steve Barley, Melissa Valentine, Woody Powell, Pam Hinds, Ece Kaynak, Tim Weiss, Aaron Horvath, Christof Brandtner, and Obaid Sarvana throughout the entirety of this project. Forrest Briscoe and three insightful reviewers were instrumental in developing this paper. This paper also benefited from feedback from Aruna Ranganathan, Matt Beane, Julia DiBenigno, Stine Grodal, Steve Vallas, Kate Kellogg, Minjae Kim, Arvind Karunakaran, Emily Truelove, Vanessa Conzon, Lucie Noury, Lindsey Cameron, and Jillian Chown. I am also grateful for seminar participants' engagement from Northwestern, MIT, Wharton, Cornell, University of Illinois, University of Michigan, Boston College, NCC, Stanford, AOM, and EGOS. Sachin Waikar and Joan Friedman provided invaluable copyediting expertise. Stanford, Northwestern, and the Psychology of Technology Institute provided funding supporting this project. Finally, this study would not have been possible without the engagement of the freelancers and clients who participated.

ORCID iD

Hatim A. Rahman https://orcid.org/0000-0002-5592-8143

Supplemental Material

Supplemental material for this article can be found in the Online Appendix at http://journals.sagepub.com/doi/suppl/10.1177/00018392211010118.

REFERENCES

Akerlof, G. A.
1970 ''The market for 'lemons': Quality uncertainty and the market mechanism.'' The Quarterly Journal of Economics, 84: 488–500.

Alkhatib, A.
2021 ''To live in their utopia: Why algorithmic systems create absurd outcomes.'' Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 8–13. New York: Association for Computing Machinery.

Alkhatib, A., and M. S. Bernstein
2019 "Street-level algorithms: A theory at the gaps between policy and decisions." Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, May. New York: ACM. DOI: 10.1145/3290605.3300760.
Allcott, H., and M. Gentzkow
2017 "Social media and fake news in the 2016 election." Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w23089.
Aloisi, A.
2015 "Commoditized workers: Case study research on labor law issues arising from a set of on-demand/gig economy platforms." Comparative Labor Law & Policy Journal, 37: 653–690.
Autor, D. H. (ed.)
2009 Studies of Labor Market Intermediation. National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press.
Barker, J. R.
1993 "Tightening the iron cage: Concertive control in self-managing teams." Administrative Science Quarterly, 38: 408–437.
Bermiss, Y. S., E. J. Zajac, and B. G. King
2014 "Under construction: How commensuration and management fashion affect corporate reputation rankings." Organization Science, 25: 591–608.
Bonet, R., P. Cappelli, and M. Hamori
2013 "Labor market intermediaries and the new paradigm for human resources." Academy of Management Annals, 7: 341–392.
Brandtner, C.
2017 "Putting the world in orders: Plurality in organizational evaluation." Sociological Theory, 35: 200–227.
Briscoe, F., and C. Murphy
2012 "Sleight of hand? Practice opacity, third-party responses, and the interorganizational diffusion of controversial practices." Administrative Science Quarterly, 57: 553–584.
Burrell, J.
2016 "How the machine 'thinks': Understanding opacity in machine learning algorithms." Big Data & Society. DOI: 10.1177/2053951715622512.
Butler, J. V., D. Serra, and G. Spagnolo
2020 "Motivating whistleblowers." Management Science, 66: 605–621.
Cameron, L.
2021 "The good bad job: Autonomy and control in the algorithmic work environment." Working paper, Wharton School.
Caplan, R., and T. Gillespie
2020 "Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy." Social Media + Society. DOI: 10.1177/205630512093663.
Cappelli, P., and M. J. Conyon
2018 "What do performance appraisals do? " ILR Review, 71: 88–116.
Castilla, E. J.
2008 "Gender, race, and meritocracy in organizational careers." American Journal of Sociology, 113: 1479–1526.
Castilla, E. J., and S. Benard
2010 "The paradox of meritocracy in organizations." Administrative Science Quarterly, 55: 543–676.
Castilla, E. J., and A. Ranganathan
2020 "The production of merit: How managers understand and apply merit in the workplace." Organization Science, 31: 909–935.

Chan, J., and J. Wang
2018 ''Hiring preferences in online labor markets: Evidence of a female hiring bias.'' Management Science, 64: 2973–2994.
Chandler, A., Jr.
1977 The Visible Hand: The Managerial Revolution in American Business. Cambridge, MA: Harvard University Press.
Charmaz, K.
2006 Constructing Grounded Theory: A Practical Guide through Qualitative Analysis, 1st ed. Thousand Oaks, CA: Sage.
Chatterji, A. K., and M. W. Toffel
2010 ''How firms respond to being rated.'' Strategic Management Journal, 31: 917–945.
Chen, L.
2016 ''Alibaba throws the spotlight on fake reviews.'' Bloomberg, Dec. 19. https://www.bloomberg.com/news/articles/2016-12-19/alibaba-lawsuit-throws-spotlight-on-brushers-gaming-rankings.
Cherry, M. A.
2015 ''Beyond misclassification: The digital transformation of work.'' Comparative Labor Law & Policy Journal, 37: 577–602.
Conger, K., and N. Scheiber
2019 ''California's contractor law stirs confusion beyond the gig economy.'' New York Times, Sept. 11. https://www.nytimes.com/2019/09/11/business/economy/uber-california-bill.html.
Curchod, C., G. Patriotta, L. Cohen, and N. Neysen
2019 ''Working for an algorithm: Power asymmetries and agency in online work settings.'' Administrative Science Quarterly, 65: 644–676.
Davis, G.
2016 The Vanishing American Corporation: Navigating the Hazards of a New Economy. Oakland, CA: Berrett-Koehler Publishers.
DiMaggio, P. J., and W. W. Powell
1983 ''The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields.'' American Sociological Review, 48: 147–160.
Dourish, P.
2016 ''Algorithms and their others: Algorithmic culture in context.'' Big Data & Society. DOI: 10.1177/2053951716665128.
Dubal, V.
2017 ''Wage slave or entrepreneur? Contesting the dualism of legal worker identities.'' California Law Review, 105: 65–123.
Edwards, R. C.
1978 ''The social relations of production at the point of production.'' Insurgent Sociologist, 8: 109–125.
Elsbach, K. D., and R. M. Kramer
1996 ''Members' responses to organizational identity threats: Encountering and countering the *Business Week* rankings.'' Administrative Science Quarterly, 41: 442–476.
Espeland, W. N.
2015 ''Noticing numbers: How quantification changes what we see and what we don't.'' In M. Kornberger, L. Justesen, A. K. Madsen, and J. Mouritsen (eds.), Making Things Valuable: 18–36. Oxford: Oxford University Press.
Espeland, W. N., and M. Sauder
2007 ''Rankings and reactivity: How public measures recreate social worlds.'' American Journal of Sociology, 113: 1–40.

Espeland, W. N., and M. Sauder
2016 Engines of Anxiety: Academic Rankings, Reputation, and Accountability. New York: Russell Sage Foundation.
Espeland, W. N., and M. L. Stevens
1998 ''Commensuration as a social process.'' Annual Review of Sociology, 24: 313–343.
Evans, J. A., G. Kunda, and S. R. Barley
2004 ''Beach time, bridge time, and billable hours: The temporal structure of technical contracting.'' Administrative Science Quarterly, 49: 1–38.
Filippas, A., J. Horton, and J. M. Golden
2019 ''Reputation inflation.'' Working paper 25857, National Bureau of Economic Research.
Finn, E.
2017 What Algorithms Want: Imagination in the Age of Computing. Cambridge, MA: MIT Press.
Fourcade, M., and K. Healy
2016 ''Seeing like a market.'' Socio-Economic Review, 15: 9–29.
Garg, N., and R. Johari
2020 ''Designing informative rating systems: Evidence from an online labor market.'' Paper presented at the 21st Conference on Economics and Computation, July 14, Association for Computing Machinery.
Gillespie, T.
2018 Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media, Illustrated Edition. New Haven, CT: Yale University Press.
Glaser, B., and A. Strauss
1999 The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine Transaction.
Gray, M. L., and S. Suri
2019 Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Boston: Houghton Mifflin Harcourt.
Heaven, W.
2020 ''This startup is using AI to give workers a 'productivity score.''' MIT Technology Review, June 4. https://www.technologyreview.com/2020/06/04/1002671/startup-ai-workers-productivity-score-bias-machine-learning-business-covid/.
Horton, J. J.
2010 ''Online labor markets.'' In A. Saberi (ed.), Internet and Network Economics: 515–522. Berlin: Springer.
Jin, G. Z., and P. Leslie
2003 ''The effect of information on product quality: Evidence from restaurant hygiene grade cards.'' Quarterly Journal of Economics, 118: 409–451.
Kailath, R.
2018 ''Some Amazon reviews are too good to be believed. They're paid for.'' NPR, July 30. https://www.npr.org/2018/07/30/629800775/some-amazon-reviews-are-too-good-to-be-believed-theyre-paid-for.
Karpik, L.
2010 Valuing the Unique: The Economics of Singularities. Princeton, NJ: Princeton University Press.
Keller, M. H.
2018 ''The flourishing business of fake YouTube views.'' New York Times, Aug. 11. https://www.nytimes.com/interactive/2018/08/11/technology/youtube-fake-view-sellers.html.
Khan, L.
2017 ''Amazon's antitrust paradox.'' Yale Law Journal., 126: 712–805.

Kittur, A., J. V. Nickerson, M. Bernstein, E. Gerber, A. Shaw, J. Zimmerman, M. Lease, and J. Horton
2013 ''The future of crowd work.'' Proceedings of the 2013 Conference on Computer Supported Cooperative Work: 1301–1318. New York: Association for Computing Machinery.
Kornberger, M., D. Pflueger, and J. Mouritsen
2017 ''Evaluative infrastructures: Accounting for platform organization.'' Accounting, Organizations and Society, 60: 79–95.
Kuhn, K. M., and A. Maleki
2017 ''Micro-entrepreneurs, dependent contractors, and instaserfs: Understanding online labor platform workforces.'' Academy of Management Perspectives, 31: 183–200.
Lambrecht, A., and C. Tucker
2019 ''Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads.'' Management Science, 65: 2966–2981.
Lamont, M.
2012 ''Toward a comparative sociology of valuation and evaluation.'' Annual Review of Sociology, 38: 201–221.
Lei, Y.-W.
2021 ''Delivering solidarity: Platform architecture and collective contention in China's platform economy.'' American Sociological Review, https://doi.org/10.1177/0003122420979980.
Leung, M. D.
2014 ''Dilettante or Renaissance person? How the order of job experiences affects hiring in an external labor market.'' American Sociological Review, 79: 136–158.
Li, J. B., and H. Piezunka
2020 ''The uniplex third: Enabling single-domain role transitions in multiplex relationships.'' Administrative Science Quarterly, 65: 314–358.
Liang, F., V. Das, N. Kostyuk, and M. M. Hussain
2018 ''Constructing a data-driven society: China's social credit system as a state surveillance infrastructure.'' Policy & Internet, 10: 415–453.
Masum, H., and M. Tovey (eds.)
2011 The Reputation Society: How Online Opinions Are Reshaping the Offline World. Cambridge, MA: MIT Press.
Morris, D.
2017 ''How an entirely fake restaurant became London's hottest reservation.'' Fortune, Dec. 10. http://fortune.com/2017/12/10/tripadvisor-london-shed-fake-restaurant/.
Oreskes, N., and E. M. Conway
2011 Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming. New York: Bloomsbury Publishing.
Orlikowski, W. J., and S. V. Scott
2014 ''What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector.'' Organization Science, 25: 868–891.
Pallais, A.
2014 ''Inefficient hiring in entry-level labor markets.'' American Economic Review, 104: 3565–3599.
Pasquale, F.
2015 The Black Box Society: The Secret Algorithms that Control Money and Information. Cambridge, MA: Harvard University Press.
Pasternack, A., and S. Melendez
2019 ''Here are the data brokers quietly buying and selling your personal information.'' Fast Company, March 2. https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information.

Pofeldt, E.
2019 ''California's new employment law has boomeranged and is starting to crush freelancers.'' CNBC, Dec. 11. https://www.cnbc.com/2019/12/11/californias-new-employment-law-is-starting-to-crush-freelancers.html.
Pope, D. G.
2009 ''Reacting to rankings: Evidence from 'America's Best Hospitals.''' Journal of Health Economics, 28: 1154–1165.
Power, M
2010 The Audit Society: Rituals of Verification. Oxford: Oxford University Press.
Proctor, R. N.
2008 ''Agnotology: A missing term to describe the cultural production of ignorance (and its study).'' In R. N. Proctor, Agnotology: The Making and Unmaking of Ignorance: 1–36. Stanford, CA: Stanford University Press.
Proctor, R. N.
2012 Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition. Berkeley, CA: University of California Press.
Rahman, H. A., and S. Barley
2016 ''Situated redesign in creative occupations—an ethnography of architects.'' Academy of Management Discoveries. DOI: 10.5465/amd.2016.0039.
Rahman, H. A., and M. A. Valentine
2021 ''How managers maintain control through collaborative repair: Evidence from platform-mediated 'gigs.''' Organization Science. DOI: 10.1287/orsc.2021.1428.
Rivera, L. A., and A. Tilcsik
2019 ''Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation.'' American Sociological Review, 84: 248–274.
Rosenblat, A.
2018 Uberland: How Algorithms Are Rewriting the Rules of Work. Oakland, CA: University of California Press.
Rosenblat, A., and L. Stark
2016 ''Algorithmic labor and information asymmetries: A case study of Uber's drivers.'' International Journal of Communication, 10: 3758–3784.
Sauder, M., and W. N. Espeland
2009 ''The discipline of rankings: Tight coupling and organizational change.'' American Sociological Review, 74: 63–82.
Scheiber, N.
2018 ''Gig economy business model dealt a blow in California ruling.'' New York Times, April 30. https://www.nytimes.com/2018/04/30/business/economy/gig-economy-ruling.html.
Schor, J. B., W. Attwood-Charles, M. Cansoy, I. Ladegaard, and R. Wengronowitz
2020 ''Dependence and precarity in the platform economy.'' Theory and Society, 49: 833–861.
Shapiro, A.
2017 ''Between autonomy and control: Strategies of arbitrage in the 'on-demand' economy.'' New Media & Society, 20: 2954–2971.
Sharkey, A. J., and P. Bromley
2015 ''Can ratings have indirect effects? Evidence from the organizational response to peers' environmental ratings.'' American Sociological Review, 80: 63–91.
Soper, S.
2020 ''Amazon drivers are hanging smartphones in trees to get more work.'' Bloomberg, Sept. 1. https://www.bloomberg.com/news/articles/2020-09-01/amazon-drivers-are-hanging-smartphones-in-trees-to-get-more-work.
Spradley, J. P.
1979 The Ethnographic Interview. New York: Holt, Rinehart and Winston.

Stark, D., and I. Pais
2021 ''Algorithmic management in the platform economy.'' Sociologica, 14: 47–72.
Strathern, M. (ed.)
2000 Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy. New York: Routledge.
Sundararajan, A.
2016 The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism. Cambridge, MA: MIT Press.
Tadelis, S.
2016 ''Reputation and feedback systems in online platform markets.'' Annual Review of Economics, 8: 321–340.
Vallas, S., and J. B. Schor
2020 ''What do platforms do? Understanding the gig economy.'' Annual Review of Sociology, 46: 273–294.
Vaughan, D.
1996 The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA. Chicago: University of Chicago Press.
Weber, M.
1978 Economy and Society: An Outline of Interpretive Sociology. Oakland, CA: University of California Press.
Weick, K. E.
1989 ''Theory construction as disciplined imagination.'' Academy of Management Review, 14: 516–531.
Zuboff, S.
2019 The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs.

Author's Biography

Hatim A. Rahman is an assistant professor of management and organizations at Northwestern University, Kellogg School of Management, 2211 Campus Dr., #5163, Evanston, IL 60208 (hatim.rahman@kellogg.northwestern.edu). His research focuses on the intersection of new technology, work, labor markets, and organizations. He earned a Ph.D. from the Center of Work, Technology, and Organization at Stanford's Management Science and Engineering.