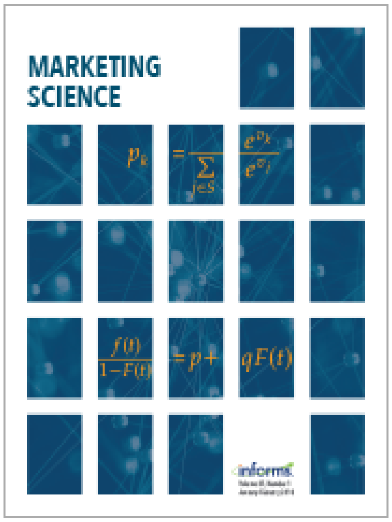
This article was downloaded by: [207.151.52.80] On: 09 January 2023, At: 10:29
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Look the Part? The Role of Profile Pictures in Online Labor Markets

Isamar Troncoso, Lan Luo

To cite this article:
Isamar Troncoso, Lan Luo (2022) Look the Part? The Role of Profile Pictures in Online Labor Markets. Marketing Science

Published online in Articles in Advance 29 Dec 2022

. https://doi.org/10.1287/mksc.2022.1425

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2022, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MARKETING SCIENCE
Articles in Advance, pp. 1–21
https://pubsonline.informs.org/journal/mksc
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Look the Part? The Role of Profile Pictures in Online Labor Markets

Isamar Troncoso,[a,]* Lan Luo[b]

[a] Department of Business Administration, Harvard Business School, Harvard University, Boston, Massachusetts 02163; [b] Department of Marketing, Marshall School of Business, University of Southern California, Los Angeles, California 90089

*Corresponding author

Contact: itroncoso@hbs.edu (IT); lluo@marshall.usc.edu, https://orcid.org/0000-0003-1002-8237 (LL)

Received: October 9, 2020
Revised: July 10, 2022; November 15, 2022
Accepted: November 15, 2022
Published Online in Articles in Advance: December 29, 2022

https://doi.org/10.1287/mksc.2022.1425

Copyright: © 2022 INFORMS

Abstract. Profile pictures are a key component of many freelancing platforms, a design choice that can impact hiring and matching outcomes. In this paper, we examine how appearance-based perceptions of a freelancer's fit for the job (i.e., whether a freelancer "looks the part" for the job), as inferred from profile pictures, can impact hiring outcomes on such platforms. Leveraging computer vision techniques and choice models, we analyze six-month data from Freelancer.com (63,014 completed jobs that received 2,042,198 applications from 160,014 freelancers) and find that, above and beyond demographics and beauty, freelancers who "look the part" are more likely to be hired. Interestingly, we do not find a strong correlation between "looking the part" and job performance. Supplementing our large-scale observational study with two choice experiments, we find that (i) the effect of perceived job fit is stronger when reputation systems are not sufficiently diagnostic to differentiate candidates and (ii) that by considering perceptions of job fit, participants are more likely to choose freelancers with fewer reviews, lower ratings, and/or without certifications. Last, we find that "platform recommendations" can only partially mitigate the unintended consequences of profile pictures, and recommending multiple freelancers can further increase the role of "looking the part."

History: Puneet Manchanda served as the senior editor for this article.
Supplemental Material: The data files and online appendices are available at https://doi.org/10.1287/mksc.2022.1425.

Keywords: freelancing platforms • profile pictures • reputation systems • platform recommendations • choice experiments • computer vision

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

1. Introduction

Freelancing platforms have gained tremendous popularity in the last decade. Such platforms allow buyers or employers (predominantly small businesses or ordinary people seeking help on tasks such as building a website, editing images, proofreading articles, etc.) and service providers or freelancers worldwide to connect and collaborate remotely. In 2019, the freelance workforce constituted 35% of the U.S. workforce.[1] In 2021, Upwork and Freelancer, the two major freelancing platforms, reported nearly $4.5 billion in gross service volume.[2] When employers list jobs on these platforms, they typically receive several applications from multiple interested freelancers. For example, Freelancer.com reports that 62% of their jobs receive applications within 60 seconds, and on average, they receive 26 applications in total. As such, deciding which freelancer would be the best fit for the job can be quite challenging in these settings.

Freelancing platforms often use different avenues to reduce information asymmetries between freelancers and employers. For instance, in addition to providing information on freelancers' skills, certifications, and past performance (e.g., the percentage of jobs completed on time and on budget), these platforms often use standardized reputation systems such as reviews and ratings (Yoganarasimhan 2013, Filippas et al. 2022). They also often allow personalized applications from freelancers where they might signal their interest and fit for the job (Ludwig et al. 2022). Furthermore, aiming to establish trust between freelancers and employers, many of these platforms actively encourage or even require freelancers to include a personal picture in their profiles. For example, in their promotional materials, Freelancer.com informs freelancers that "using your real face lets employers know that they are dealing with a real person and not a stranger."[3] Similarly, Upwork.com tells freelancers that "clients want to feel they can trust a freelancer before they engage them for a project and your profile photo is an important part of the equation."[4] Nevertheless, it is not immediately evident what role profile pictures serve in the presence of other platform design choices, such as using standardized reputation systems and platform recommendations.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Within this context, a primary goal of our research is to investigate whether and how the presence of profile pictures, as encouraged or mandated by many freelancing platforms, may impact hiring outcomes. Specifically, we explore the extent to which employers' perceptions of a freelancer's job fit (whether a freelancer is *perceived* to have the skills and abilities required for the job), as inferred from a profile picture, can affect which freelancer gets hired. Moreover, we conjecture and empirically explore two conceptual underpinning of such perceptions. First, perceptions of job fit may be formed holistically, based on multiple visual cues in a picture, many of which could extend beyond demographics and beauty. For example, visual cues such as wearing glasses can influence perceptions of intelligence (Wei and Stillwell 2017), which may well influence perceptions of fit for jobs associated with that trait (e.g., jobs in the fields of science, technology, engineering, and mathematics). Second, perceptions of job fit could be job category specific, that is, the same profile picture might be perceived differently depending on the job under consideration. For instance, although a "nerdy-looking" profile picture (demographics and beauty held constant) may be perceived as a high fit for a programming task, it may be perceived as a low fit for creative marketing tasks.

Our research is inspired by anecdotal evidence suggesting that people often rely on appearances to judge whether an individual is suitable for a certain kind of job. One such example is the social media backlash to OneLogin's recruiting ads in 2015, which according to some users, failed to represent what a female engineer should look like, partly because the woman portrayed in one of the ads was considered too attractive to be a real engineer.[5] Many other stories on the web illustrate that the stereotypical image of a software engineer is "normally male White or Asian who is average height and build, not athletic and prefers little to no human contact."[6] Research has shown that appearance-based judgments can have downstream consequences and impact decisions (Olivola and Todorov 2010b). In the political domain, for example, research has shown that appearances-based perceptions of candidates' competence can predict election outcomes better than chance (Olivola and Todorov 2010a) and that merely having a "conservative-looking" face seems to benefit candidates running in conservative areas (Olivola et al. 2012). To the best of our knowledge, however, no academic research has yet investigated the potential downstream consequences of appearance-based perceptions on hiring decisions in online labor marketplaces.

In the marketing literature, Luo et al. (2008) suggested that consumers often use both objective and subjective criteria to evaluate products. In a similar vein, we explore whether employers in freelancing platforms use both arguably more objective criteria (e.g., online reputation variables, performance metrics, and certifications) and arguably more subjective criteria (e.g., whether the freelancer "looks the part" for the focal job, which we later find contains very little information about a freelancer's actual work quality) when deciding whom to hire. To this end, we perform a large-scale observational study and two experimental studies to seek answers to the following three research questions:

- What role does perceived freelancer-job fit, as inferred from profile pictures, play in hiring outcomes on freelancing platforms?
- What is the interplay between perceptions of freelancer-job fit and reputation systems in hiring choices?
- Can "platform recommendations" help reduce the role of appearance-based perceptions of freelancer-job fit in hiring choices?

Our observational study is based on Freelancer.com, one of the largest freelancing platforms worldwide. We analyzed data on 63,014 jobs (which received 2,042,198 applications from 163,014 freelancers) completed between January and June 2018 in the two major categories on the platform (website, information technology (IT), and software, and design, media, and architecture). We leverage modern computer vision techniques (Hartmann et al. 2021, Zhang et al. 2021, Zhang and Luo 2022) to capture perceptions of freelancers' job fit as inferred from their profile pictures and use interpretable machine learning methods to explore the drivers of such perceptions. We find that what makes a picture "look the part" depends on the job category: for programmers, demographics, facial features (gender, race, and smile), and certain backgrounds and accessories (computer in view, wearing glasses) are the most important predictors of high perceived fit, whereas for designers, the esthetic quality of the image matters most, and demographics are less important predictors of high perceived fit. We then use choice models to estimate hiring decisions and find that freelancers who are perceived as a higher fit for the job (i.e., those who "look the part") are more likely to be hired, even after controlling for demographics and beauty. This effect is small in magnitude, but comparable to (i) a change of 5% in the asking price, a difference that could help cover the 10% fee freelancers pay to the platform; and (ii) a change of 0.3 stars in the average rating, a reputation variable that freelancers might have less control over than their profile pictures. We also find that the effect of "looking the part" is stronger when the pool of applicants for the focal job is more similar, suggesting that it could serve as an additional differentiation tool to help freelancers secure jobs in such highly competitive markets.[7] Interestingly, we also note that perceptions of job fit do not strongly correlate with the rating that freelancers receive after completing the job nor whether the job is completed on time, suggesting

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

that such perceptions might not be strong signals of freelancers' quality.

Nevertheless, our findings from the observational study might be confounded by supply biases (e.g., freelancers may strategically decide which job to apply for or what price to charge depending on whether they "look the part") and can only be interpreted as correlational rather than causal. To address these concerns, we conduct two choice experiments where we randomize freelancers' characteristics in each choice set to explore the causal effect of perceived job fit and leverage several manipulations to examine the role of different platform design choices that we cannot readily accomplish using secondary data.

In our first experimental study, we implement a 2×2 between-participants choice experiment based on choice-based conjoint studies (Luo et al. 2008, Aribarg et al. 2017). We use a factorial design to orthogonalize freelancers' perceived job fit from gender and race and separate the effect of "looking the part" from such variables and manipulate the diagnosticity of the reputation system and the availability of profile pictures. We find that perceived job fit has a causal impact on hiring choices above and beyond gender and race and that the effect of perceived job fit is stronger when the reputation system is less diagnostic. The effect of perceived job fit is more salient for freelancers who pass the consideration stage, and disclosing profile pictures can significantly hurt the choice share of freelancers with sufficiently desirable attributes on aspects other than their pictures to pass the consideration stage but "do not look the part." Last, we find that when profile pictures are available on the platform, participants take into account perceptions of job fit and choose a greater percentage of freelancers with fewer reviews, lower average ratings, and/or without certifications. Thus, our findings suggest that the use of profile pictures on freelancing platforms can lead to some unintended consequences, such as hurting the hiring prospects of freelancers who "do not look the part" and inadvertently encouraging employers to de-emphasize arguably less-noisy metrics of freelancers' quality.

In our second experimental study, we implement a 3×1 between-participants choice experiment to explore whether "platform recommendations" can help weaken the unintended consequences of disclosing freelancers' profile pictures. Our focus on such a design choice is motivated by their wide adoption (e.g., Amazon's choice, editor recommendations such as Apple's "New Apps we love," etc.) and the observed variation in their design and implementation in practice (e.g., Freelancer.com recommends one freelancer per job, Upwork.com recommends multiple freelancers per job, and Fiverr.com recommends no freelancers). Our findings suggest that platform recommendations might partially help mitigate the unintended consequences of profile pictures by making participants more likely to choose freelancers with a high reputation. However, recommending one freelancer does not seem to help freelancers who have a high reputation but "do not look the part," and recommending multiple freelancers can further increase the effect of perceived job fit.

Our research makes the following contributions to the literature. First, we contribute to the literature on the role of profile pictures in online marketplaces (Pope and Sydnor 2011, Doleac and Stein 2013, Edelman and Luca 2014, Ert et al. 2016, Hannák et al. 2017) by showing that profile pictures can influence hiring outcomes based on appearance-based perceptions above and beyond well-studied prejudice variables such as demographics or beauty (i.e., in our context, whether a freelancer "looks the part" for a job). To the best of our knowledge, our research is the first empirical study evincing such a role of profile pictures in online labor marketplaces. Second, we contribute to the literature on online marketplaces and the role of reputation systems (e.g., Sun 2012, Yoganarasimhan 2013, Watson et al. 2018, Benson et al. 2020) and platform recommendations (e.g., Senecal and Nantel 2004, Gai and Klesse 2019, Kawaguchi et al. 2019, Liang et al. 2019, Bairathi et al. 2022) by carefully examining the interplay between such design choices and the role of profile pictures. In particular, we show that "looking the part" plays a greater role in final choices when reputation variables become less diagnostic and when the platform recommends multiple freelancers per job. We believe that this study is among the first to examine and uncover the intertwined role of profile pictures with these other important platform design choices.

Our findings offer valuable insights for both freelancing platforms and freelancers. For platforms, our findings underscore the potential unintended consequences of disclosing profile pictures. Moreover, we show that to weaken the role of profile pictures, platforms should explore alternative methods to make their reputation systems more diagnostic, that is, provide enough variation for employers to differentiate freelancers more easily. We also show that platforms should design their recommendations carefully to avoid the unintended effects of recommending multiple freelancers, which can further strengthen the role of profile pictures. For freelancers, our findings provide evidence that they might better position themselves in highly competitive marketplaces by making their profile pictures "look the part." Although perceptions of fit are partially explained by characteristics that freelancers cannot change (such as gender and race), these perceptions are also influenced by visual cues that they can control, such as accessories, background, and image quality.

The remainder of the paper is organized as follows. We summarize the related literature in Section 2, describe our observational study in Section 3, present our first and second experimental studies in Sections 4 and 5, respectively, and conclude the paper in Section 6.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

2. Related Literature

Our work relates to three streams of literature. The first stream explores the role of profile pictures in online marketplaces. Prior research has shown that the use of profile pictures can facilitate gender and racial biases in peer-to-peer lending markets (Pope and Sydnor 2011), e-commerce (Doleac and Stein 2013), and freelancing platforms (Hannák et al. 2017, Leung et al. 2020). Some researchers have also investigated the role of visual cues beyond demographics and beauty, such as how smile and body prominence can affect property demand on Airbnb (Zhang et al. 2020) or lending rates (Athey et al. 2022). In this paper, we extend this stream of literature by exploring the potential role of *perceived job fit*, a subjective judgment of whether the freelancer looks suitable for a certain type of job (whether the freelancer "looks the part"). Specifically, we emphasize that such perceptions of job fit (i) may be formed holistically based on the conjunction of multiple visual cues that can go above and beyond well-studied visual cues such as demographics, beauty, and smile and (ii) are job category specific (e.g., an image that is perceived as a high fit as a programmer may not be perceived as a high fit as a graphic designer).

The second stream of literature relates to the role of reputation systems in online marketplaces. Within the context of online labor marketplaces, prior research has shown that employers and freelancers rely significantly on reputation variables to make their decisions (Yoganarasimhan 2013, Benson et al. 2020) and that past rating on similar jobs are good predictors of future performance (Kokkodis and Ipeirotis 2016). Nevertheless, ratings are highly top-censored, and most freelancers have perfect scores (Filippas et al. 2022), a phenomenon that reduces the informativeness of reputation systems. Drawing from accessibility-diagnosticity theories of consumer behavior (Feldman and Lynch 1988, Lynch et al. 1988) and empirical findings that the influence of reputation variables hinges on their diagnosticity (Watson et al. 2018), we contend that when reputation systems are less diagnostic, employers are more likely to rely on perceptions of freelancers' job fit. In other words, such perceptions might become an additional dimension to differentiate among highly rated freelancers. To our knowledge, our research is among the first to explore the interplay between profile pictures and reputation systems in hiring outcomes.

The third stream of literature relates to the role of platform recommendations in online marketplaces. Prior research has shown that platform recommendations have a positive impact on choices of the featured products (Senecal and Nantel 2004, Benlian et al. 2012, Bairathi et al. 2022) and related products (Liang et al. 2019), the role of the framing of the recommendation (e.g., "People who like this also like" versus "Similar to this item"; Gai and Klesse 2019), and contextual factors (e.g., time and crowd pressure; Kawaguchi et al. 2019). In this paper, we are interested in exploring whether platform recommendations (Adomavicius et al. 2011) can also help reduce the role of profile pictures in hiring choices. To our knowledge, our research is among the first to explore such an interplay between profile pictures and platform recommendations in online labor marketplaces.

3. Observational Study

In this section, we present an observational study in which we explore (i) what makes an image to be perceived as a high fit for jobs in different categories, (ii) the relationship between perceived job fit and hiring outcomes, and (iii) the relationship between perceived job fit and job performance.

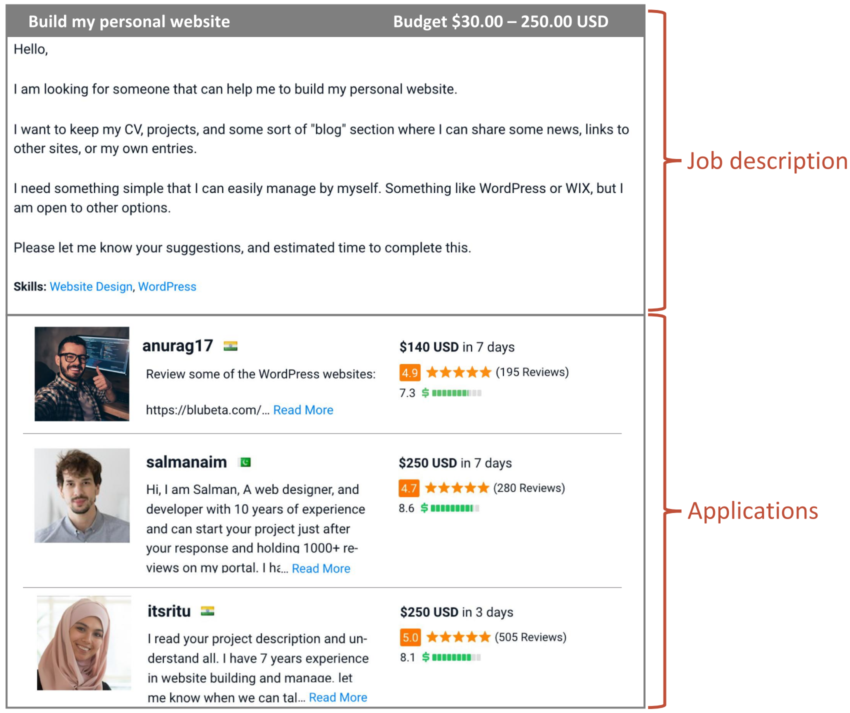
3.1. Data

Our observational study uses data from Freelancer.com, the world's largest freelancing platform as measured by the number of users and jobs. To join the platform, freelancers must register and create their user profiles, including a description of the services they provide, their skills, a summary of their expertise, and their profile pictures. To hire a freelancer, employers need to register, post their jobs, and wait for freelancers to apply for them.[8] In Figure 1, we illustrate a job with 3 of the more than 70 applications it received within one day. As shown in the upper panel of Figure 1, a job posted by the employer includes a description of the task, the budget, and desired freelancer skills. As shown in the bottom panel of Figure 1, applications submitted by freelancers include the requested price, a summary of the freelancer's reputation variables (number of reviews and average rating), and an application description (text), among other details. After the employer decides whom to hire and the freelancer completes the job, both have the option to review each other.

Using the website application programming interface (API), we collect data for all the jobs posted between January and June 2018. To obtain granular insights on perceptions of job fit at the job category level, we focus on jobs within the two largest job categories on the platform, which together account for nearly 67% of all job listings: (i) website, IT, and software and (ii) design, media, and architecture (43% and 24% of the job listings, respectively). Our final sample consists of 63,014 jobs that ended with a successful hire and the 2,042,198 applications they received from 160,014 different freelancers.[9] In the following, we provide descriptive statistics on freelancers' reputation variables and the number of applications and characteristics of the applicants a job receives.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Figure 1. (Color online) Example of a Job Posted by the Employer and Applications Submitted by Freelancers



Notes. We replace the original pictures with licensed images purchased from an online stock photography company called Shutterstock for illustration purposes. The perceived job fit based on these pictures exhibits much less variation from actual freelancer profile pictures available at Freelancer.com.

3.1.1. Freelancers' Reputation. Among freelancers who have received at least one review, the average number of reviews is 33, with an average rating of 4.7 of 5 stars. Moreover, at the freelancer level, the median average rating is 5 stars, illustrating an extremely positive distribution of ratings typical of these types of platforms (Filippas et al. 2022).

3.1.2. Number of Applications and Applicants' Characteristics at the Job Level. On average, jobs in the websites, IT, and software category receive 27 applications, 11 of which are submitted by freelancers with a reputation above average (at least 33 reviews and 4.7 average rating). Similarly, jobs in the design, media, and architecture category receive an average of 39 applications, of which 20 are submitted by freelancers with an above average reputation.

Overall, these statistics suggest that online labor marketplaces are quite competitive: It is not unusual for employers to have a large pool of highly rated candidates to inspect and for freelancers to compete against.

3.2. Creating Profile Picture–Related Variables

Approximately 94% of the freelancers in our sample have a profile picture. Here, we describe how we use these pictures to infer four variables that prior literature has shown affect hiring outcomes: gender, race, age, and beauty. Then, we explain how we use these pictures to create our primary variable of interest: perceived job fit. Last, we examine what makes a profile picture "look the part" and to what extent these perceptions of fit are explained by gender, race, and beauty.

3.2.1. Inferring Gender, Race, Age, and Beauty. We use the Face++ computer vision API to identify whether a profile picture contains a human face (instead of logos or avatars) and, if so, the apparent gender, race, age, and beauty of the freelancer. To validate these labels, we recruit human raters to manually code a subsample of images and find a high level of agreement with the labels provided by the API (see Online Appendix C for more details). We note that 72% of the profile pictures contain a human face (28% are logos or avatars). Within this group, most profile pictures are labeled as male freelancers (79% male, 21% female). In terms of freelancers' race, the majority of the profile pictures are labeled as Indian (49%) and White (28%), followed by Black (14%) and Far East Asian (9%).[10] In terms of age, the average and standard deviation are 29 and 8, respectively. Last, we observe that in terms of the beauty score, which ranges from zero to one,

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

the average and standard deviation are 0.630 and 0.115, respectively.

3.2.2. Labeling Profile Pictures Based on Perceived Job Fit. A critical step in preparing the data for our analyses is to label profile pictures based on *perceived job fit*, that is, a label to approximate whether, based on a freelancer's profile picture, the employer perceives the freelancer to have sufficient skills to meet the demands of a specific job. Given our need to create such labels for 163,014 freelancers in each of the two job categories (326,028 labels in total), it is prohibitive to manually label all profile pictures in our observational data using human raters. Therefore, we leverage modern computer vision techniques to overcome this challenge. Consistent with our conjecture that perceptions of job fit could result from a complex weighting of multiple visual patterns in an image, we use a deep learning image classifier to generate such labels. In the following, we summarize the three underlying steps in this process and refer readers to Online Appendix D for more details.

The first step is to create a training set. For this purpose, we select a random subsample of 3,000 profile pictures and recruit raters with experience in each job category to score them based on their perceptions of the freelancers' fit for a job in each job category. We approximate how employers (mostly small business owners or individuals who might have some but often not extensive hiring experiences) on such platforms might perceive freelancers' profile pictures by recruiting one group of raters with experience in the software and IT industry and another group with experience in the graphic design industry (see details in Online Appendix D). Each image is rated by three to five domain-specific raters using a five-point scale, wherein 1 represents the lowest possible job fit and 5 represents the highest possible job fit. The interrater agreement for the perceived job fit as a programmer and as a designer, as measured by Cronbach's alpha, is 0.764 and 0.696, respectively. Following standard practices in the literature, we take the average scores across raters for each job category and convert the continuous measure into binary levels (low and high perceived fit) to mitigate potential noise in the training data (Zhang et al. 2015, 2021; Liu et al. 2019; Zhang and Luo 2022). As a result of this process, each image has two labels: high/low perceived job fit as a programmer and high/low perceived job fit as a graphic designer. We find a modest correlation of 0.22 between the two, consistent with our claim that such perceptions are job category specific.

The second step is to train the classifier. For this purpose, we use 80% of the labeled images to train the model and 20% to validate its performance. We follow three standard practices to reduce overfitting problems that can arise due to the modest size of our training set: transfer learning (Hartmann et al. 2021, Zhang et al. 2021, Zhang and Luo 2022), data augmentation techniques (Krizhevsky et al. 2017), and dropout techniques (Srivastava et al. 2014). We use different convolutional neural networks (CNN) architectures, including VGG16, Inception, and ResNet. The architecture that performs the best in our setting is VGG16, which achieves an out-of-sample area under the receiver operating characteristic curve of 88.89% and 85.84% for the perceived job fit as a programmer and as a graphic designer, respectively. We provide more details about the architecture, training process, and performance metrics in Online Appendix D.

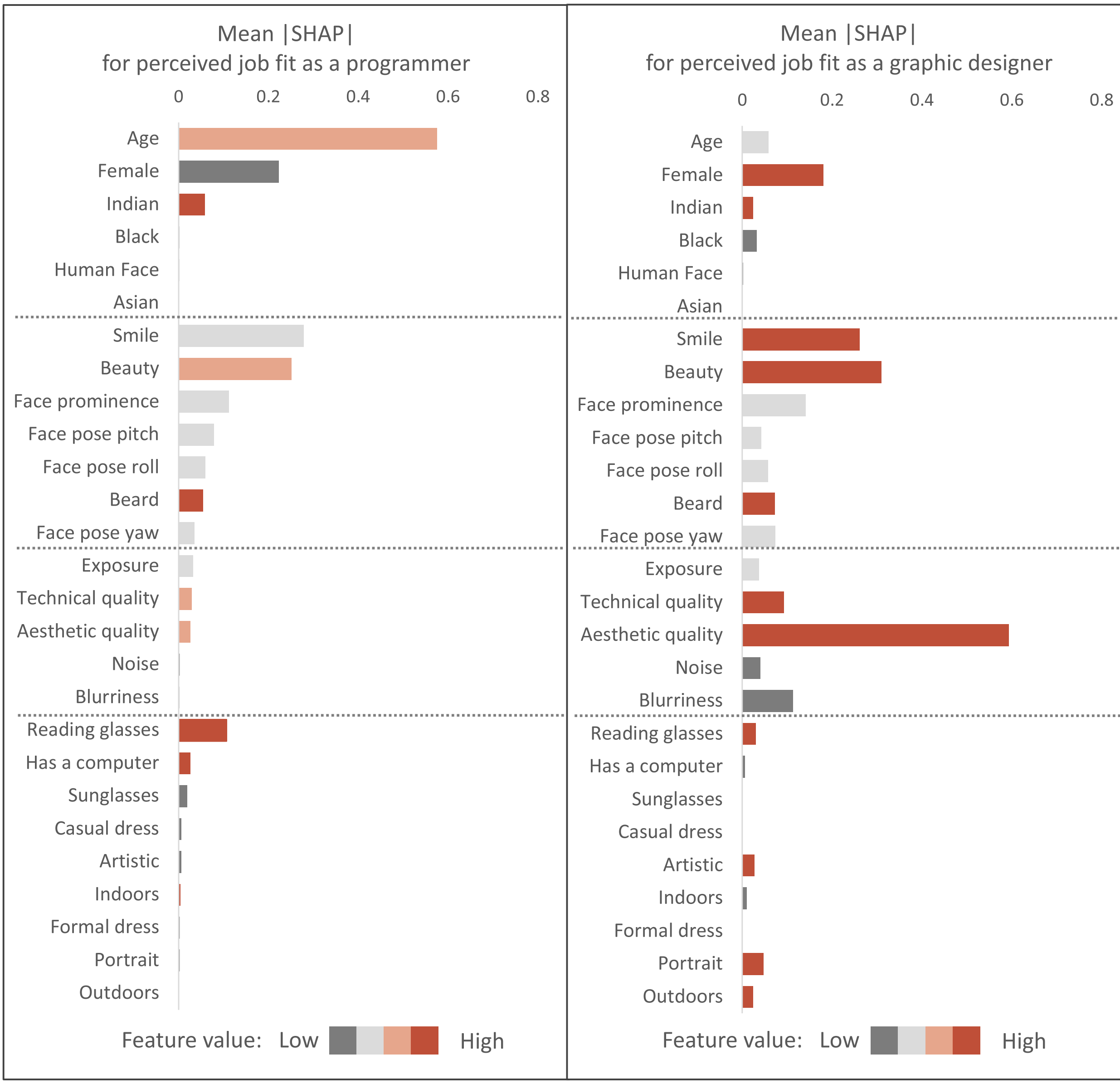
The third and final step is to use the parameters learned by the classifier to generate a job category–specific perceived job fit label for all the profile pictures in our data. Specifically, each image receives two labels corresponding to the predicted probability of being classified as a high perceived job fit for each job category. In Online Appendix D, we show that these labels are robust to alternatives in the training process (i.e., freezing the last convolutional block to reduce the number of trainable parameters) and alternative definitions of the score (i.e., using the input to the classification layer rather than the predicted probability).

3.2.3. What Makes a Profile Picture "Look the Part". Although our deep learning image classifier allows us to extrapolate human ratings with relatively high accuracy, it provides little insight into what makes an image "look the part" for a specific job category. In this section, we use interpretable machine learning methods to address such a limitation.

We start by collecting interpretable image-related features using different computer vision APIs (see Figure 2 for the complete list of features included in our analysis and Online Appendix E for a detailed description of them). We broadly categorize these features into four groups: (i) demographic variables (e.g., gender and race); (ii) facial features (e.g., smile and beauty); (iii) image quality (e.g., exposure and blurriness); and (iv) accessories and background (e.g., reading glasses and computer). These image-related features are included in our analysis either because of prior research (Fagerstrøm et al. 2017, Wei and Stillwell 2017, Lennan et al. 2018) or our conjecture that they might play a role in perceived job fit.

Next, we adopt the Xgboost classifier (Chen and Guestrin 2016) to predict whether an image is labeled as a "high perceived job fit" for each category based on human ratings (training set described in Section 3.2.2) as a function of the features listed previously. For interpretability, we use Shapley values (Lundberg et al. 2020) to explore what image-related features are more important in predicting such labels. We summarize our main results in Figure 2, which illustrates the

Figure 2. (Color online) Shapley Value of Different Image-Related Features



Notes. The dashed lines separate features from different groups, namely demographic variables, facial features, image quality, and accessories and background. Features are ordered from highest to lowest important for each group in the programmer label. Feature order is fixed in both plots to facilitate the comparison. Color codes indicate the relationship between feature values and high perceived job fit; for example, middle/high levels of age are associated with higher perceived fit as a programmer, whereas low levels of age are associated with higher perceived fit as a graphic designer.

feature importance based on Shapley additive explanations (SHAP), measured as the average of the absolute SHAP values, and uses color code to indicate the relationship between feature values and high perceived job fit in each category. We refer the reader to Online Appendix E for more details about the training process, including the selection of hyperparameters and model performance.

The results in Figure 2 suggest that, although well-known prejudice variables (demographics and beauty) play an important role in predicting perceptions of job fit, other image-related attributes such as accessories, background, and image quality also help explain such perceptions. To further support these findings, in Online Appendix F, we present another study showing that modifying the background and accessories on an image can make the same freelancer to be perceived as a lower/higher fit for the job. Moreover, we find that the role of different features varies with the job category. For example, although image quality plays a prominent role in predicting perceived job fit for graphic designers, it is negligible in predicting perceived job fit for programmers. Even features of comparable importance, such as gender

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

and beauty, play different roles for different categories. For example, being a female is associated with lower perceptions of fit as a programmer but with higher perceptions of fit as a designer. Similarly, higher levels of beauty are considerably more important in perceptions of fit as a designer than those of a programmer.

Overall, the findings support our conjecture that perceptions of job fit can go above and beyond well-known prejudice variables (gender, race, and beauty) and that they are job category specific. Henceforward, we use the perceived job fit score as predicted by the image classifier in our analyses, as it is a single measure that summarizes the job-specific role of different image features and might capture the role of other interpretable and noninterpretable image features that cannot be obtained using computer vision API.

3.3. How Does "Looking the Part" Relate to Hiring Choices?

In our setting, each job is characterized by an employer hiring a freelancer from a choice set that is unique in terms of both (i) the number of freelancer applicants and (ii) who apply for the job. To model hiring decisions in this setting (similar to Chan and Wang 2018), we use a conditional logit model (McFadden 1973), where employer i chooses the freelancer j in the pool of applicants for job t to maximize his/her utility u_{ijt}, specified as

$$\begin{aligned} u_{ijt} = \alpha_t &+ \beta_1 \mathbf{ProfilePicture}_j + \beta_2 \mathbf{Reputation}_j \\ &+ \beta_3 \mathbf{Application}_j + \beta_4 \mathbf{Performance}_{jt} \\ &+ \beta_5 \mathbf{Controls}_{ij} + \varepsilon_{ijt}, \end{aligned} \tag{1}$$

where ε_{ijt} is a stochastic component that follows a Gumbel distribution. The independent variables in Equation (1) include the following (see detailed explanation and descriptive statistics of these variables in Online Appendix G):

- *Profile Picture:* Set of variables extracted from the freelancer's profile picture, including whether there is a profile picture, and if so, the perceived job fit for that job category, and whether there is a human in the picture, and if so, the apparent gender, race, age, and beauty of that freelancer.
- *Reputation:* Set of reputation variables, including whether the freelancer has no reviews yet, the number of reviews, average rating, average sentiment score (valence), and average sentiment magnitude (strength) extracted from the text of the reviews.
- *Application:* Set of variables that describe the application submitted by a freelancer (see bottom part of Figure 1), including the price and number of days the freelancer requests to complete the job, the log of the word count in the text of the application, the similarity between the job application and the job description, the distance to the prototypical text application in the focal job category, whether the freelancer is recommended by the platform for the focal job (highlighted in the top position as seen by the employer), and the relative position of the application in the list displayed to the employer.
- *Performance:* Set of variables that describe a freelancer's performance on previous jobs, including the total earnings made on similar jobs, the percentage of prior jobs completed on time, and the percentage of prior jobs completed on budget.
- *Additional Controls:* Set of additional controls including whether the freelancer has a certification, whether the freelancer has passed exams on skills required by the employer, the freelancer's region of residence, whether the freelancer is from a developed country, whether the freelancer is from the same country as the employer, whether the employer has reviewed the freelancer in the past, among others.

To explore the role of perceived job fit on hiring choices, we estimate different specifications of Equation (1), and report the estimated coefficients in Table 1. In column 1, we consider a baseline model that does not include any profile picture-related variables. The coefficients for the reputation and application variables are consistent with prior findings in the literature (Yoganarasimhan 2013, Kokkodis et al. 2015, Chan and Wang 2018). For example, as the number of reviews and the average rating increases, a freelancer's probability of being hired also increases. Conversely, as the price increases, a freelancer's probability of being hired decreases. We also observe that the probability of being hired increases with the job application word count and the similarity between the employer's description and the freelancer's application and decreases with the prototypicality of the job application in that category (i.e., less prototypical job applications might stand out). Henceforward, we consider this specification a benchmark because it includes all possible information employers can obtain without a freelancer's profile picture.

In column 2, we incorporate our primary variable of interest, perceived job fit,[11] which has a positive and significant coefficient ($\beta = 0.089, p < 0.01$), suggesting that freelancers who are perceived as a higher fit for the job are more likely to be hired.[12] The model fit, as measured by the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), favors this specification relative to the baseline model (column 1), suggesting that perceived job fit can help explain the observed hiring outcomes. Last, in column 3, we incorporate well-known prejudice variables (demographics and beauty) as additional controls. Under this specification, the perceived job fit coefficient is still positive and significant but shrinks slightly, which is consistent with our findings that such perceptions of job fit are partially but not fully explained by gender, race, age, and beauty.

Thus far, we have shown that "looking the part" is positively correlated with the probability of being

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Table 1. Estimating Hiring Choices in Observational Data

	(1)	(2)	(3)
Profile pictures variables			
Perceived Job Fit Score		0.089***	0.075***
Has Picture		0.283***	0.287***
Reputation variables			
No Reviews Yet	−0.668***	−0.684***	−0.706***
Log(1 + N. Reviews)	0.437***	0.438***	0.435***
Avg. Rating	0.312***	0.308***	0.306***
Application variables			
Offered Price	−1.865***	−1.868***	−1.869***
Log(1 + Application WC)	0.161***	0.161***	0.162***
Application-Description Similarity	1.292***	1.289***	1.286***
Distance to Prototypical Application	0.678***	0.681***	0.684***
Recommended by the Platform	0.297***	0.297***	0.298***
Additional variables			
Other Application Variables	✓	✓	✓
Control Variables	✓	✓	✓
Human		✓	✓
Demographics (gender, race, age)			✓
Beauty			✓
N	2,028,764	2,028,764	2,028,764
LL	−154,866	−154,839	−154,788
AIC	309,828	309,780	309,690
BIC	310,430	310,419	310,403

Notes. Conditional logit estimates with standard errors clustered at the job level. The dependent variable is whether employer i hired freelancer j from the pool of applicants for job t.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

hired.[13,14] However, how meaningful is this relationship? To quantify the relative impact of "looking the part," we use coefficients in column 2 of Table 1 to estimate how a one-standard-deviation (SD) change in perceived job fit, average rating, number of reviews, or a 5% change in the offered price would impact hiring choices. We report these results in Table 2. In column 2, we show the percentage of times employers would switch their choice if we were to change one characteristic of the hired freelancer. We find that if the perceived job fit of the hired freelancers were 1 SD lower (0.351 points lower), employers would hire a different freelancer 3.6% of the time. Although the magnitude of this effect is much smaller than that from four fewer reviews, it is comparable to a 1-SD decrease in the average review rating and a 5% increase in the offered price. Although freelancers could improve their profile pictures at a relatively low cost (see Online Appendix F for more evidence on this), reputation variables require considerable time and effort to change (freelancers need to both get hired and receive positive reviews). In column 3 of Table 2, we conduct a similar exercise but change the characteristics of the second-best candidates; thus, those with the second-highest choice probability based on the estimates in column 2 of Table 3. These results suggest that if their perceived job fit were 1 SD higher (0.351 points higher), these freelancers would get 6.5% more jobs. Again, the effect is much smaller than that of the number of reviews, but it is comparable to the effect of average ratings and a 5% change in prices.

Moreover, we contend that a relatively small effect of perceived job fit could still make a difference,

Table 2. Quantifying the Magnitude of the Effect of Different Variables on Hiring Choices

		Percentage switching when variable is	
	Absolute change (1)	Worse for winner (2)	Better for second best (3)
One-SD change in perceived job fit	0.351 points	3.572%	6.533%
One-SD change in average rating	0.266 stars	4.964%	9.631%
One-SD change in log(1 + N of reviews)	4 reviews	14.621%	73.844%
5% change in offered price	$ 7	4.228%	10.790%

Notes. Column 1 shows the absolute value associated with a one-standard-deviation (SD) change for each variable and a 5% change in the average transaction price. Column 2 shows the percentage of times we predict employers would switch their choice if their chosen freelancer had a different value in the focal attribute. Column 3 shows the percentage of times we predict employers would switch their choice if the second-best freelancer (second-highest choice probability) had a higher value in the corresponding attribute. Predictions are based on the hiring parameters in column 2 of Table 1.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

perhaps serving as an additional differentiation tool in such a competitive market where many highly rated freelancers apply for the same job. To formally explore this conjecture, we examine whether the main effect of perceived job fit (column 2 of Table 1) strengthens as the competitiveness of the applicant pool for the focal job increases. Our first proxy for competitiveness measures the dispersion among the reputation variables of the applicants within each job. Specifically, we compute the interquartile range of (i) the number of reviews and (ii) the average rating of the applicants and explore their interactions with the perceived job fit variable. A larger (smaller) interquartile range means more (less) dispersion in the reputation variables of the pool of applicants. As shown in column 2 of Table 3, the relationship between perceived job fit and hiring outcomes becomes stronger when the pool of applicants is more similar in terms of their number of reviews (i.e., when the dispersion of the number of reviews decreases). We find no significant interaction between perceived job fit and dispersion of the average ratings (column 1 of Table 3), which we believe could result from the small variation in average ratings on the platform.

Our second proxy for competitiveness measures the similarity of the applicants for each job. Specifically, we take a subset of characteristics that are salient in the applications (Figure 1) and that have a significant impact on hiring choices, X = {number of reviews, average rating, total earnings for jobs in the category, price, certification, relevant qualification, from a developed country, from employer's country}, and compute the Euclidean distance between each applicant in the pool and the hired freelancer. As shown in column 3 of Table 3, the perceived job fit coefficient decreases when the distance between the hired freelancer and the second closest candidate increases. Similarly, in column 4 of Table 3, we find that the perceived job fit coefficient decreases when the average distance between the hired freelancer and the five closest candidates increases. In other words, these results suggest that the relationship between perceived job fit and hiring outcomes strengthens as the similarity among the pool of applicants increases (i.e., when the distance among them decreases).

3.4. Are Freelancers Who "Look the Part" Better at Their Jobs?

Thus far, we have shown that employers are more likely to hire freelancers who are perceived as a higher fit for the job based on their profile pictures. A potential follow-up question is whether such perceptions are merely subjective or informative. In other words, could

Table 3. How Does the Perceived Job Fit Coefficient Vary with the Competitiveness of the Choice Set?

	(1)	(2)	(3)	(4)
Profile pictures variables				
Has Picture	0.280***	0.277***	0.303***	0.306***
Perceived Job Fit Score	0.076***	0.133***	0.212***	0.261***
⋯ × *Dispersion in Avg. Rating*	0.010			
⋯ × *Dispersion in N. Reviews*		−0.031***		
⋯ × *Distance to next best candidate*			−0.153***	
⋯ × *Avg. Distance to 5 next best candidates*				−0.138***
Reputation variables				
No Reviews Yet	−0.683***	−0.681***	−0.653***	−0.654***
Log(1 + N. Reviews)	0.438***	0.438***	0.438***	0.438***
Avg. Rating	0.309***	0.309***	0.315***	0.315***
Application variables				
Offered Price	−1.868***	−1.868***	−1.875***	−1.875***
Log(1 + Application WC)	0.161***	0.161***	0.162***	0.162***
Application-Description Similarity	1.290***	1.293***	1.286***	1.286***
Distance to Prototypical Application	0.681***	0.680***	0.682***	0.683***
Recommended by the Platform	0.298***	0.298***	0.299***	0.299***
Additional variables				
Performance variables	✓	✓	✓	✓
Other application variables	✓	✓	✓	✓
Control variables	✓	✓	✓	✓
Human	✓	✓	✓	✓
N	2,028,764	2,028,764	2,028,010	2,028,010
LL	−154,838	−154,835	−154,661	−154,653
AIC	309,780	309,775	309,426	309,410
BIC	310,431	310,426	310,077	310,061

Notes. Conditional logit estimates with standard errors clustered at the job level. The dependent variable is whether employer i hired freelancer j from the pool of applicants for job t.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

freelancers who "look the part" be higher-quality workers? In this section, we explore this possibility by examining the relationship between perceived job fit and proxies for freelancers' quality, such as job performance and qualifications.

First, we look at the correlations between perceived job fit and freelancers' (i) qualification metrics, (ii) reputation variables, and (iii) performance metrics at the beginning of our sample period. We report our results in Table 4. Overall, the correlation coefficients range from zero to 3%, indicating either negligible or very small (albeit statistically significant in some cases) correlations between perceived job fit and these freelancer quality proxies.[15]

Second, we regress the perceived job fit on (i) whether the freelancer receives a review upon completing a job; (ii) whether the job is completed on time and on budget; and (iii) the ratings and the sentiment of the review. After controlling for all information employers observe when making a hiring decision, we observe no significant positive correlation between perceived job fit and any outcome metric (see Online Appendix H.4 for more details). Taken at face value, these findings suggest that, although perceptions of freelancer job fit are positively related to hiring choices, such perceptions are rather noisy signals of the freelancer quality that have little additional information above and beyond reputation and performance metrics.

The findings presented in this section are subject to some limitations common to observational studies. Given that freelancers might endogenously decide which job to apply for and which price to request, we can only interpret our results as correlational rather than causal. In the experimental studies we present in the following sections, we address these concerns by employing an orthogonal design to randomize freelancer and application characteristics within each choice set. Furthermore, the lack of variability in the observational study compromises our ability to fully answer our research quesions. First, because ratings are extremely positive throughout the platform, we cannot explore the interplay between the role of profile pictures and the diagnosticity of the reputation system. We address this limitation by manipulating the diagnosticity of the reputation system in our first experimental study (Section 4). Second, because we only observe one platform design, we cannot explore how different platform designs impact hiring outcomes. We address this limitation by manipulating the availability of profile pictures in our first experimental study (Section 4) and the platform recommendations in our second experimental study (Section 5).

Table 4. Exploring the Correlation Between Freelancers' Perceived Job Fit Score and Other Performance, Certification, and Reputation Variables

	Freelancers with	
	At least one review	No reviews yet
Number of qualifications	−0.030***	0.009***
Average score in qualifications	0.001	0.016***
Log(1 + N reviews)	−0.007	
Average rating	0.020***	
Percentage of jobs on completed on time	0.012**	
Percentage of jobs on completed on budget	0.008	

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

4. Experimental Study 1

In this section, we present a choice experiment in which we explore (i) the interplay between the role of profile pictures and the diagnosticity of the reputation system, (ii) the role of profile pictures in different stages of the decision process (consideration versus choice), and (iii) the impact of disclosing freelancers' profile pictures on hiring outcomes.

4.1. Experimental Design

Our first experimental study is a choice-based conjoint experiment. Participants were asked to imagine that they would like to hire a programmer from a freelancing platform to build a personal website. We chose this scenario under the following three considerations: (i) it falls under the largest job category on the platform we use for our observational study; (ii) personal websites can serve multiple purposes, some of which could be relevant for a broad pool of participants; and (iii) we can easily familiarize participants with the desired outcome with some illustrative examples. After being presented with the scenario, participants were given 14 choice tasks, each presenting a hiring scenario with 10 hypothetical freelancers as the applicants. For each hiring scenario, participants are asked to indicate all the freelancers that they would consider hiring and, among those they would consider, the one freelancer they would prefer to hire.

The experiment is a 2 × 2 between-participants design, in which we manipulate (i) the diagnosticity of the reputation system and (ii) the availability of profile pictures. Here, we explain the details of our experimental design.

4.1.1. Choice Profiles. Although our observational study is highly information rich with a large number of controls, a typical conjoint experiment often does not comprise more than six or eight attributes (Orme 2002, Luo et al. 2008, Aribarg et al. 2017). As such, we identify six attributes (perceived job fit, gender, race, online reputation, price, and certification) based on (i) the research questions we want to explore and (ii) their importance in hiring outcomes as discovered from our observational study. Following the convention in conjoint

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

studies, we instructed the participants to assume that all attributes not presented are constant across all job applicants (Rao 2014).

We use profile pictures to depict the first three attributes: perceived job fit, gender, and race. To separate out the effects of race and gender from that of perceived job fit, we incorporate 16 different combinations of perceived job fit (low, high), race (Far East Asian, Black, Indian, and White), and gender (female, male). Namely, we include both low and high job fit pictures for each race-gender combination so that perceived job fit is balanced and orthogonal to race and gender variables by design. This setup allows us to address endogeneity concerns related to supply biases and correct for other imbalances in our observational data (e.g., disproportionate participation of freelancers from one race). Furthermore, we include four profile pictures per possible combination of perceived job fit, gender, and race to ensure that our findings are not subject to idiosyncratic characteristics related to one individual picture. The levels of perceived job fit are defined by human raters, that is, as in the training data used in Section 3.2.2, and we keep the same image size used by the platform in our observational study (64 × 64 pixels), so that our choice experiment can mimic that platform to the greatest extent possible.

The fourth attribute is reputation, which we defined as the combination of the number of reviews and average rating, and we manipulated as explained in more detail later. The fifth attribute is price. To capture the realistic price range for this type of job on the platform, we listed a similar job on Freelancer.com and scanned similar jobs in our secondary data. Subsequently, we defined price levels as \$100, \$150, and \$200. The last attribute is certification, which indicates whether the freelancer has successfully passed an exam on a relevant skill for the job.[16]

Based on the six attributes described previously (also summarized in Table 5), we use an orthogonal factorial design to generate 14 choice tasks with 10 alternatives each, including two hold-out tasks that are presented at the beginning of the survey. Although most conjoint studies do not include this many alternatives, we aim to mimic the number of applications employers can simultaneously compare on most freelancer platforms. For example, Freelancer.com splits the total number of applications into multiple pages and displays eight applications per page, whereas Upwork.com displays 10 applications per page. It is also worth noting that this number is significantly smaller than the total number of applicants a job receives on these platforms (on average, at least 27 applications in our observational data). Therefore, our choice tasks are a simplified version of the actual process most employers go through in practice and allow us to strike a balance between avoiding overwhelming participants with too many alternatives and approximating a realistic scenario to the greatest extent possible.

4.1.2. Manipulating the Diagnosticity of the Reputation System. We manipulate the levels of the reputation attribute to create two diagnosticity conditions: (i) a less diagnostic and (ii) a more diagnostic condition. To define the reputation levels for each condition, we post a job similar to that used in our choice experiment on Freelancer.com (i.e., recruiting a freelancer to create a personal website) and use the statistics of the applicants for that job as reference values (Table 6). We illustrate these two conditions in Figure 3.

For the less diagnostic condition, we use four reputation levels resulting from the combination N reviews = {30, 300} × average rating = {4.8, 5.0} values that correspond to the 25th and 75th percentiles of the reference values in Table 6. Considering that we only show 10 applications per choice task and that, on average, jobs in this category receive at least 11 applications from freelancers with more than 30 reviews and an average rating above 4.7, we believe that this condition provides a reasonable representation of the platform.

Table 5. Attribute and Respective Levels Used in the First-Choice Experiment

Attribute	Description	Levels
Perceived job fit	Whether the candidate is perceived as a high fit for the job, as inferred from his/her profile picture	Low, high
Gender	Apparent gender of the candidate, as inferred from the profile picture	Female, male
Race	Apparent race of the candidate, as inferred from the profile picture	Asian, Black, Indian, White
Reputation	Number of reviews and average rating (N reviews, Avg. Rating)	Less diagnosticity condition: (30, 4.8), (300, 4.8), (30, 5.0), (300, 5.0) More diagnosticity condition: (10, 4.5), (300, 4.5), (10, 5.0), (300, 5.0)
Price	Requested price to complete the job	\$100, \$150, \$200
Certification	Whether the candidate has certified a skill that is relevant for the job	Yes, no

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

For the more diagnostic condition, we use four reputation levels resulting from the combination N reviews = {10, 300} × average rating = {4.5, 5.0}. At first glance, these numbers might look similar to those in the less diagnostic condition. They are, however, considerably more diagnostic than (i) the reference levels in Table 6, (ii) the values we observe in the observational data, and (iii) the levels observed in online labor platforms in general (Filippas et al. 2022). Moreover, prior research has also shown that small differences in the number and average rating of the reviews can significantly impact consumer preferences and choices for products (Watson et al. 2018). We also ran a pretest and found that, as intended, participants perceived freelancers to be less similar in the more diagnostic condition than in the less diagnostic condition (see details in Online Appendix I)

4.1.3. Manipulating the Availability of Profile Pictures. We manipulate the availability of profile pictures to create two conditions: (i) with and (ii) without profile pictures. The first condition resembles the current platform design, where employers can immediately see freelancers' profile pictures along with their reputation variables. The second condition is used as a benchmark and is motivated by traditional labor markers, where recruiters are not allowed to request the inclusion of photographs in job applications.[17] Moreover, we use the second condition as a baseline to quantify the impact of disclosing profile pictures on hiring outcomes.

4.1.4. Data Collection and Incentives. We recruited 399 college students from a major U.S. university. The participants self-selected into taking part in our study with a lottery-based reward. We informed participants that we would randomly select two lottery winners, for whom we would hire a programmer to build them a personal website based on their hiring preferences (as inferred from their responses to the survey). The winner would also receive an Amazon gift card for $210 minus the cost of hiring the programmer, an additional incentive we include to ensure participants do not ignore the price attribute when making their decisions. We randomly assigned participants to one of the four experimental conditions. To control for order effects, we randomize the order in which the freelancers' applications are shown to participants in each choice task.

4.2. Results

We separately estimate hiring choices for each experimental condition using a hierarchical Bayesian multinomial logit. We report the estimated population average utilities and attribute importance weights in Table 7. The signs of reputation, price, and certification parameters are consistent with our findings from the observational data and prior literature. Thus, ceteris paribus, participants prefer freelancers with higher reputation levels, who offer a low price, and have a certification. To further evaluate the performance of the model, we compute out-of-sample hit rates using the two hold-out tasks. Considering that our choice tasks have ten applicants per choice task, we compare our results to a baseline of a 10% hit rate (i.e., prediction by chance). The minimum out-of-sample hit rate was 38% across the four conditions, that is, 3.8 times better than the baseline. We summarize our main findings here.

4.2.1. Effect of Perceived Job Fit on Hiring Choices. We now explore the role of perceived job fit on hiring decisions by focusing on the results obtained under the "with picture condition" (columns 1 and 2 of Table 7). The first result worth highlighting is that, at the population level, perceived job fit has a positive and significant effect on hiring choices after separating out the effect of gender and race, supporting the idea that the effect of perceived job fit can go above and beyond demographic variables.

By making the reputation system more diagnostic, the importance weight of this attribute increases, and the importance weights of most of the other attributes decrease. More importantly, using the less diagnostic condition as a baseline, we note that in the more diagnostic condition, the relative decrease in the importance weight of perceived job fit (28.536% change) is more pronounced than those of price (7.506% change) and certification (12.612% change) and more pronounced than the relative increase in the importance weight of reputation (12.503% change). In a similar vein, we find that the willingness to pay for a freelancer with a picture that "looks the part" is higher

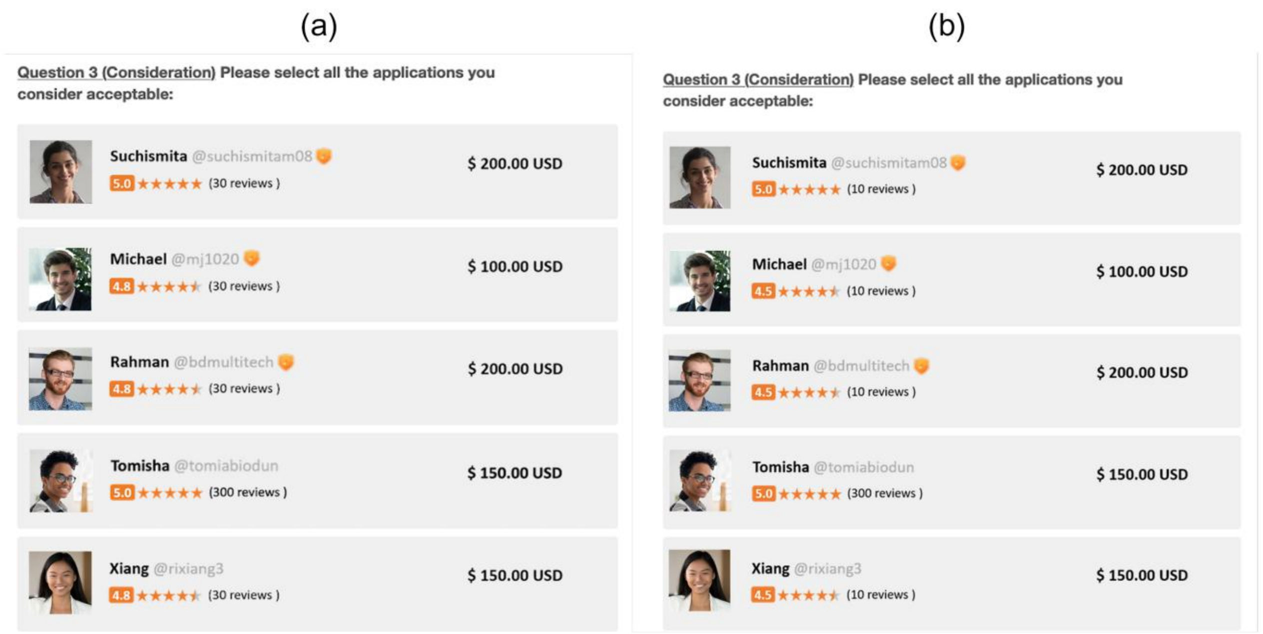
Table 6. Summary Statistics of Applications Received for a Job Listed in Freelancer.com to Recruit a Freelancer to Create a Personal Website (Hiring Scenario Used in Experimental Studies)

	Minimum	25th percentile	50th percentile	75th percentile	Maximum
Number of reviews	1	27	88	309	1,992
Average rating	4.3	4.8	4.9	5.0	5.0
Price	$30	$120	$150	$220	$450

Notes. We received a total of 75 applications for this job. The number of reviews and average ratings were obtained for the subset of 86% freelancers that have at least one review.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Figure 3. (Color online) Experimental Study 1: Manipulating the Diagnosticity of the Reputation System



Notes. (a) Less diagnostic condition. (b) More diagnostic condition. Example of 5 of 10 candidates in one conjoint card across different diagnosticity conditions. All alternatives, attributes, and pictures are fixed across conditions except for the numerical values (number of reviews and ratings) used to represent the reputation levels. We replace the original pictures with licensed images purchased from an online stock photography company called Shutterstock for illustration purposes. The perceived job fit based on these pictures exhibits much less variation from actual freelancer profile pictures available at Freelancer.com.

when the reputation system is less diagnostic ($14 versus $11). Also, ignoring perceived job fit decreases the out-of-sample hit rate by roughly three percentage points only when the reputation system is less diagnostic. Overall, these findings suggest that the effect of perceived job fit is more pronounced when the reputation system is less diagnostic, which is consistent with our conjecture that profile pictures can serve as an additional differentiator, especially in highly competitive settings.

4.2.2. Role of Perceived Job Fit in Consideration vs. Choice. We now examine the role of perceived job fit in different stages of the decision process. We are especially interested in exploring whether participants put more emphasis on profile pictures early in the decision process when they form a consideration set or later when they need to make a hiring choice among those they would consider.

Recall that participants were asked to select all freelancers they would consider before making their choice. We note that when the reputation system is less diagnostic, on average, participants select 4.4 freelancers (out of a total of 10) in the consideration stage of each hiring scenario. When the reputation system is more diagnostic, this number decreases to 3.9, and the difference across conditions is statistically different from zero ($p < 0.01$).

To explore the role of perceived job fit in the consideration versus choice stages of the decision-making process, we compare the percentage of times that freelancers who "look the part" are considered and chosen for the job. As shown in Figure 4, freelancers who are perceived as a high fit for the job are more likely to be considered (left panel) and to be chosen (right panel) than those who are perceived as a low job fit. Nevertheless, the disparity between the two groups is considerably more pronounced in the choice stage than in the consideration stage of the decision process. Taken at face value, these results suggest that, relative to the minor screening role profile pictures might play in earlier stages of the decision process, they play a more pronounced differentiator role in later stages of the decision process.

4.2.3. Impact of Disclosing Profile Pictures on Hiring Outcomes. We further explore the impact of profile pictures on hiring outcomes using responses from the "without profile pictures" condition as a baseline. First, to explore the impact of disclosing profile pictures from the freelancers' perspective, we measure the shift in freelancers' choice shares as the difference between the choice share in the "with picture condition" minus the choice share in the "without picture condition." Then, for each diagnosticity condition, we examine the shift in choice shares of the top freelancer

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

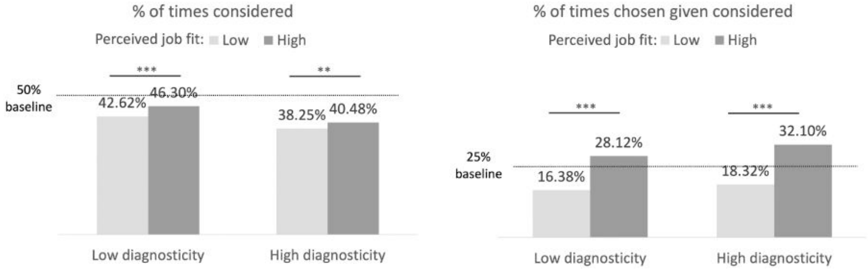
Table 7. Average Utilities and Attribute Importance in Experimental Study 1

	With picture		Without picture	
	Less diagnostic (1)	More diagnostic (2)	Less diagnostic (3)	More diagnostic (4)
Average utilities				
Perceived job fit				
High	0.844*** (0.871)	0.561*** (0.556)		
Reputation				
High ratings × High N reviews	7.573*** (2.661)	8.204*** (2.491)	7.313*** (1.774)	10.574*** (3.474)
High ratings × Low N reviews	3.051*** (0.772)	2.029*** (0.877)	2.418*** (1.179)	4.327*** (1.635)
Low ratings × High N reviews	4.682*** (4.682)	4.997*** (1.494)	4.238*** (1.093)	6.815*** (2.678)
Price				
Low	5.646*** (2.728)	4.974*** (2.756)	5.390*** (2.447)	5.331*** (2.880)
Mid	3.207*** (1.516)	3.664*** (2.137)	3.014*** (1.102)	3.580*** (1.656)
Certification				
Yes	4.054*** (2.321)	3.266*** (1.531)	3.661*** (1.600)	4.434*** (2.156)
Additional controls				
Gender	✓	✓		
Race	✓	✓		
Average attribute importance (%)				
Perceived job fit	5.302	3.789		
Gender	2.841	4.929		
Race	7.709	6.989		
Reputation	37.179	41.827	45.561	51.661
Price	27.807	25.720	32.294	27.249
Certification	19.162	16.745	22.144	21.090

Notes. Reference levels are "Low" for perceived job fit, "Low ratings: Low *N* reviews" for reputation, "High" for price, and "No" for certifications. Estimates of average utilities include the population posterior mean and standard deviation of the heterogeneity distribution in parentheses.
***$p < 0.01$.

of each choice task, that is, freelancers who have the highest choice share in the "without picture" condition. Aggregating such shifts across the choice tasks, we note that when the reputation system is less diagnostic, there is a significant difference between the shift in choice shares for freelancers who "look the part" and those who "do not look the part" ($p < 0.05$), with the former staying almost the same and the latter group losing an average of 10 percentage points in their choice shares due to picture disclosure. When we inspect this result more carefully, we note that such a decrease is explained by participants switching to the second or third top freelancer, who have sufficiently high reputation variables and a more favorable profile picture. When the reputation system is more diagnostic, freelancers are perceived as less similar in terms of their reputation variables, and we observe no significant differences in the shift in choice shares. Overall, these findings suggest that when the reputation systems are less diagnostic, disclosing profile pictures can significantly hurt the hiring prospects of freelancers who rank higher on aspects other than their pictures but "do not look the part."

Figure 4. (Color online) Role of Perceived Job Fit on Consideration vs. Choice



Notes. (Left) Dotted line indicates a 50% baseline equal to the probability of being considered by chance. (Right) Dotted line indicates a 25% baseline equal to the probability of being chosen by chance, taking into account the fact that participants select an average of four freelancers in the consideration stage. Significance levels: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

To explore the impact of disclosing profile pictures from the platform's perspective, we examine the distribution (%) of the attribute levels among the chosen freelancers under different experimental conditions and report these results in Table 8. When the reputation system is less diagnostic, participants in the "with picture" condition have a greater tendency to choose freelancers with lower ratings, lower number of reviews, or both than those in the "without picture" condition. Similarly, when the reputation system is more diagnostic, participants in the with picture condition are more likely to choose freelancers with lower ratings and less likely to choose freelancers with certifications. Thus, when profile pictures are disclosed, participants place more emphasis on subjective perceptions of freelancers' fit for the job in the decision-making process and less weight on (arguably) more-objective proxies of freelancers' quality (such as the number of reviews, rating, and certification). Overall, these findings suggest that profile pictures in online labor marketplaces might negatively impact matching outcomes.

5. Experimental Study 2

In this section, we present a choice experiment in which we explore the interplay between the role of profile pictures and another platform design choice that has gained popularity in many online marketplaces: "platform recommendations." Platform recommendations (e.g., Amazon's choice) aim to help with the evaluation of products/services on the platform, and they can significantly impact user choices (e.g., Senecal and Nantel 2004, Gai and Klesse 2019, Kawaguchi et al. 2019, Bairathi et al. 2022). As such, the goal of this experiment is to explore whether recommending freelancers based on their reputation and credentials can help mitigate the unintended consequences of profile pictures.

In practice, we observe significant variation in the design and implementation of platform recommendations. For instance, some freelancing platforms (such as Freelancer.com and PeoplePerHour.com) feature only one freelancer per job, others like Upwork.com feature multiple freelancers per job (three to four recommendations, as per our experience with this platform), and others like Fiverr.com recommends no freelancers at all. In this study, we explore whether and how such differences might play a role in their effectiveness in reducing the role of "looking the part" in hiring outcomes.

5.1. Experimental Design

Our second experimental study is a choice-based conjoint experiment, similar to that presented in Section 4.1, and we use the same scenario and choice tasks described there. Furthermore, we start from the experimental condition that resembles the characteristics of the platform used in the observational data, that is, the condition with profile pictures and a less diagnostic reputation system. From there, we manipulate the number of platform recommendations as explained here.

5.1.1. Manipulating the Number of Platform Recommendations. We manipulate the number of platform recommendations to create three conditions: (i) no recommendations, (ii) one recommendation, and (iii) three recommendations. These conditions are chosen to represent the variation we observe in the field. The first condition corresponds to the same "with picture condition" used in the first experimental study. In the last two conditions, we use a "recommended freelancer" badge to feature the one or three freelancers with the highest reputation levels in each hiring scenario, and inform participants that "the platform might recommend one (or multiple) freelancers who, based on his/her previous jobs and ratings, could be a good fit for your job." This criterion is motivated by those observed on these platforms and in conversations with industry experts. Should this criterion result in ties, that is, that two or more freelancers have the same reputation levels, we recommend the freelancer with certification when possible or randomly select

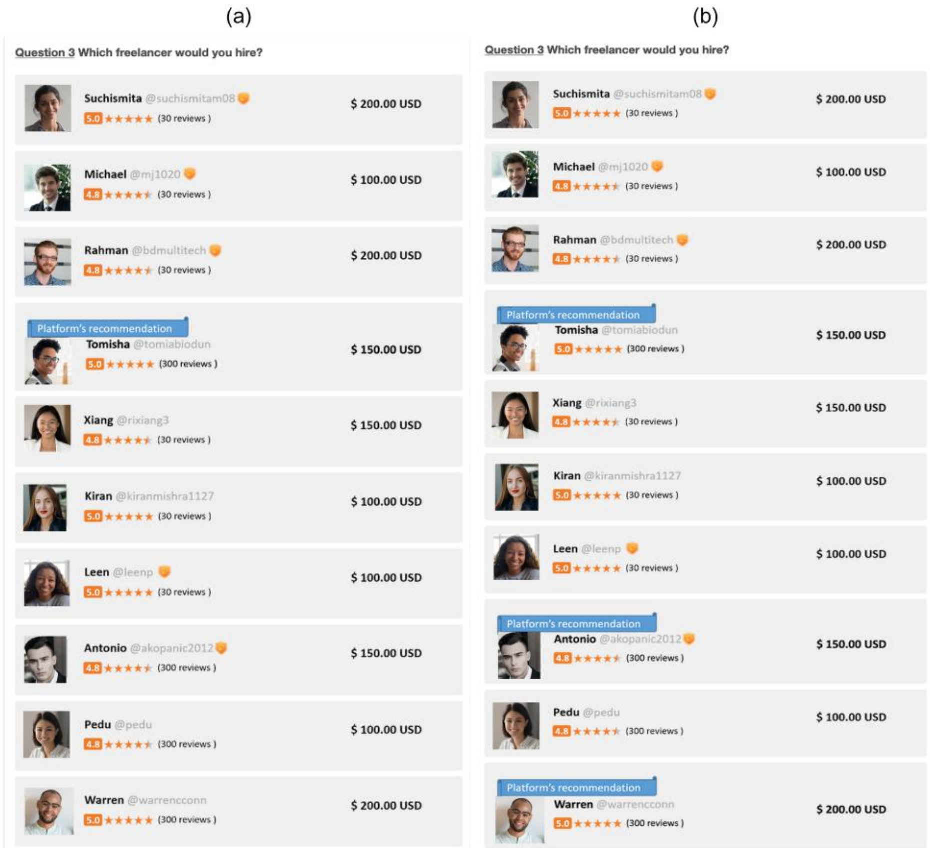
Table 8. Percentage Distribution (%) of Attribute Levels of the Chosen Freelancers

		Low diagnosticity		High diagnosticity	
		No picture	Picture	No picture	Picture
Reputation	*High rating × High N reviews*	**72.109**	**65.512**	**75.327**	**75.765**
	High rating × Low N reviews	**8.503**	**9.241**	**6.291**	**3.316**
	Low rating × High N reviews	**17.177**	**21.040**	**17.320**	**18.963**
	Low rating × Low N reviews	**2.211**	**4.208**	**1.062**	**1.956**
Price	*Low*	61.650	63.944	54.167	54.167
	Mid	33.163	30.776	38.889	38.520
	High	5.187	5.281	6.944	7.313
Certification	*Yes*	85.459	84.158	**86.846**	**82.058**
	No	14.541	15.842	**13.154**	**17.942**

Notes. Percentage of the chosen freelancers with a certain attribute-level combination. Thus, for a given attribute, the numbers sum up to 100. For each diagnosticity condition and attribute, we highlight in bold the distribution of levels that are significantly different in the picture and the no picture condition ($p < 0.01$ based on the chi-squared test).

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Figure 5. (Color online) Experimental Study 2: Manipulating the Number of Recommendations



Notes. (a) One recommendation. (b) Three recommendations. Example of one conjoint card across different platform recommendations conditions. All alternatives and attributes are fixed across conditions, except for the number of freelancers being recommended by the platform. We replace the original pictures with licensed images purchased from an online stock photography company called Shutterstock for illustration purposes. The perceived job fit based on these pictures exhibits much less variation from actual freelancer profile pictures available at Freelancer.com.

one of the freelancers with the same attributes. We illustrate the last two conditions in Figure 5.

5.1.2. Data Collection and Incentives. We recruited 425 college students from the same university as in our first experimental study (none were allowed to participate in both studies), along with a lottery-based reward. The description of the reward and the data collection process are the same as those described in Section 4.1. We randomly assign participants to one of the three experimental conditions. To control for order effects, we randomize the order in which the freelancers' applications are shown to participants in each choice task.

5.2. Results

We separately estimate hiring choices for each experimental condition using a hierarchical Bayesian multinomial logit model. We report the estimated population average utilities and attribute importance weights in Table 9.

We find that platform recommendations indeed significantly impact participant choices. Because recommended freelancers have high reputation levels, recommendations help improve matching outcomes (at least from the platform perspective). We further carry out pairwise t tests to quantify changes in the perceived job fit importance among participants across the three conditions. We find that, relative to the no recommendation condition, the attribute importance of perceived job fit decreases slightly when recommending one candidate (0.263 percentage points or 6.808%), but such a difference is not statistically different from zero. More interestingly, we discover that recommending multiple freelancers significantly increases the role of profile pictures on hiring choices (1.082 percentage points or 28.051% relative to the no recommendation condition and 1.345 percentage points or 37.41% relative to the one recommendation condition).

In sum, our results suggest that platform recommendations can partially mitigate the unintended consequences of profile pictures. Because recommendations significantly impact choices, recommending freelancers with a high reputation makes participants more likely to choose freelancers of arguably higher quality, hence improving matching outcomes from the platform's perspective. However, from the freelancers' perspective, being the

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Table 9. Average Utilities and Attribute Importance in Experimental Study 2

	Number of recommendations		
	None (1)	One (2)	Three (3)
Average utilities			
Perceived job fit			
High	0.438*** (0.523)	0.571*** (0.573)	0.827*** (0.784)
Reputation			
High ratings: High N reviews	6.605*** (2.374)	6.258*** (1.496)	5.903*** (2.468)
High ratings: Low N reviews	1.945*** (1.226)	1.760*** (0.721)	2.321*** (1.412)
Low ratings: High N reviews	4.045*** (1.797)	3.636*** (3.636)	3.499*** (1.477)
Price			
Low	4.516*** (2.277)	6.025*** (2.359)	5.345*** (2.993)
Mid	2.776*** (1.430)	3.273*** (1.248)	3.193*** (1.825)
Certification			
Yes	3.156*** (2.103)	2.273*** (1.832)	2.207*** (1.966)
Recommendation			
Yes		2.537*** (1.334)	2.811*** (1.626)
Additional controls			
Gender	✓	✓	✓
Race	✓	✓	✓
Average attribute importance (%)			
Perceived job fit	3.859	3.596	4.941
Gender	2.701	3.287	4.066
Race	8.071	5.565	7.054
Reputation	40.783	32.223	30.065
Price	26.663	30.419	27.316
Certification	17.924	12.112	11.576
Recommendation		12.797	14.981

Notes. Reference levels are "Low" for perceived job fit, "Low ratings: Low *N* reviews" for reputation, "High" for price, "No" for certifications, and "No" for recommendations. Estimates of average utilities include the population posterior mean and standard deviation of the heterogeneity distribution in parentheses.
***$p < 0.01$.

solo recommended freelancer does not seem to help reduce the disadvantage that freelancers with high reputation levels but "do not look the part" face and being one of the multiple recommended freelancers can further exacerbate such a disadvantage.

6. Conclusions

Freelancing platforms have gained tremendous popularity in the last decade, facilitating the matching of millions of employers and freelancers worldwide. These online labor marketplaces are quite competitive, with employers usually having a large pool of highly rated candidates to inspect and freelancers having a long list of competitors applying for the same position. This paper empirically explores how freelancers' profile pictures can impact hiring outcomes in such a competitive setting. Using an observational study, we show that freelancers who "look the part" are more like to be hired. Such an effect goes above and beyond demographics and beauty, and its magnitude is comparable to (i) a change of 5% in the asking price, a difference that could help freelancers cover the 10% fee they must pay to the platform; and (ii) a change of 0.3 stars in the average rating, a reputation variable that could be hard to change in the short term. We also implement two experimental studies to further explore the interplay between the "look the part" effect and different platform design choices. We find that perceived job fit becomes more important for hiring decisions (i) when reputation systems are less diagnostic and (ii) in later stages of the decision-making process, where they seem to serve as an additional means to differentiate highly competitive candidates who pass the consideration stage. Using a hypothetical platform design without profile pictures as a benchmark, we also find that profile pictures decrease the matching quality: By taking into account perceptions of job fit, participants choose a higher percentage of freelancers with fewer reviews, lower average ratings, and/or without certifications. Last, we find that, although platform recommendations might partially help with mitigating the unintended consequences of profile pictures by making participants more likely to choose freelancers of (arguably) higher quality, recommending multiple freelancers can further increase the role of perceptions of job fit on choices.

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

Our research makes the following contributions to the literature. First, we contribute to the literature on the role of profile pictures in online marketplaces (e.g., Pope and Sydnor 2011, Doleac and Stein 2013, Edelman and Luca 2014, Ert et al. 2016, Hannák et al. 2017). Specifically, we show that, above and beyond well-studied prejudice variables such as demographics or beauty, profile pictures can influence hiring decisions based on appearance-based perceptions of the focal candidate's fit for the job (whether the candidate "looks the part"). Second, we contribute to the literature on online marketplaces and the role of reputation systems (e.g., Sun 2012, Yoganarasimhan 2013, Watson et al. 2018, Benson et al. 2020) and platform recommendations on choices (e.g., Senecal and Nantel 2004, Gai and Klesse 2019, Kawaguchi et al. 2019, Liang et al. 2019). Specifically, we show that the role of profile pictures intertwines with such platform design choices and that "looking the part" matters more for hiring decisions when reputation systems are less diagnostic and when platforms recommend multiple freelancers for the job.

Our findings also have important implications for platforms. In particular, we suggest that encouraging or mandating profile pictures can hurt the hiring prospects of freelancers who "do not look the part" and might inadvertently make employers deemphasize arguably less noisy signals of freelancers' quality (e.g., reputation and performance variables). We also show that having a more diagnostic reputation system and reducing the number of freelancers they recommend can help mitigate such unintended consequences of disclosing profile pictures. Our findings also have implications for freelancers. In particular, we discover that certain cues over which freelancers can easily control, such as accessories, background, and image quality, can help elicit a more positive perception of their fit for a job and better position themselves in such highly competitive marketplaces.

Our research also provides some fruitful directions for future work. First, despite our best efforts to recruit a relevant pool of participants and use lottery-based incentives, the external validity of the findings from our experimental studies is still limited. Future research could collaborate with freelancing platforms to run field experiments inspired by our experimental designs. Second, although we focus on the role of profile pictures on hiring outcomes, future research could explore the role of videos. For example, Fiverr.com allows freelancers to upload a video to their profiles to present themselves to potential employers. A video may have an even stronger impact than a profile picture on employers' perceptions of a freelancer, for instance, through the dynamics of their body language, especially facial gestures (e.g., dynamics of a smile; Krumhuber et al. 2009) or through their voice (e.g., perceptions of competence; Burgoon 1978). Third, future research can further explore alternative platform design solutions to reduce the unintended consequences of profile pictures. Despite our multiple attempts, we have yet to discover an effective avenue to mitigate such effects other than increasing the diagnosticity of reputation systems, which can be challenging to implement in practice. Last, although we focus on the context of online labor marketplaces, profile pictures are a key design element of many other platforms, including some that only facilitate the initial touch point between employers and service providers (e.g., LinkedIn.com; Care.com; Healthgrades.com). Future studies could extend our research by exploring the business implications of digital profile pictures and appearance-based inferences in these alternative settings.

Acknowledgments

The authors thank participants at the Carnegie Mellon University Conference on Digital Marketing and Machine Learning and the INFORMS Marketing Science Conference and seminar attendants at Columbia Business School, Georgetown University, Harvard Business School, Nanyang Technology University, New York University, Rice University, Temple University, University of Arizona, University of British Columbia, University of California at Berkeley, University of Maryland, University of Miami, University of Notre Dame, University of Pittsburgh, University of Texas at Dallas, University of Wisconsin-Madison, and Washington University in St. Louis for helpful comments. The authors also thank Sawtooth Software for supporting this research.

Endnotes

[1] Source: https://www.upwork.com/i/freelancing-in-america/2019/.

[2] Source: Presentations and reports publicly available on https://investors.upwork.com/ and https://www.freelancer.com/investor#ReportsAndPresentations.

[3] Source: https://www.freelancer.com/community/articles/profile-picture-tips-and-tricks.

[4] Source: https://support.upwork.com/hc/en-us/articles/211063208-Sample-Profiles-and-Best-Practices.

[5] Source: https://medium.com/the-coffeelicious/you-may-have-seen-my-face-on-bart-8b9561003e0f.

[6] See https://www.linkedin.com/pulse/armando-you-dont-look-like-software-engineer-armando-pantoja/.

[7] Some may wonder about the possibility of freelancers using fake profile pictures. If such were the case, employers should respond little to this type of "cheap talk" (Pope and Sydnor 2011), and we should find no effect of profile pictures. In other words, if fake profile pictures induce a possible confound in our estimates, it should go against what we aim to uncover, and our observational study's findings would provide a lower bound of the effect of interest. Additionally, platforms' efforts to verify freelancers' identity and impose penalties on users who use fake profile pictures may further relieve these concerns.

[8] Employers also have the option to search for freelancers and directly invite them for the focal job. We analyze these outcomes in Online Appendix B and find the main result is consistent with that presented in this section (i.e., that freelancers who "looking the part" are more likely to be chosen).

[9] We provide additional details on jobs that ended without a hire in Online Appendix A. Typically, jobs that end without a hire tend to be unusual (e.g., very high budget) or from employers who might not have serious intentions to hire (e.g., employers did not verify their payment method or make a deposit).

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

[10] The race labels provided by the Face++ API do not include a Hispanic label. Although prior research has recognized the difficulty of distinguishing between White and Hispanic race when inferring race from profile photos (Davis et al. 2019), we further address this concern using alternative race labels provided by Clarifai API, which include American Indian, Asian, Black, Hawaiian or Pacific Island, Hispanic, Middle Eastern, and White. Because a large number of freelancers in our data are from Asia, we opted for the race variable based on the Face++ API, which distinguishes Indians from Far Eastern Asians.

[11] This model specification implicitly assumes that there are no measurement errors in the perceived job fit, that is, that employers observe the same perceived job fit score as predicted by our image classifier. Although our image classifier does not have perfect accuracy, its errors are fairly balanced (see Online Appendix D, Table A5), and measurement errors around the perceived job fit variable are likely to be noise. In Online Appendix H.1, we show that the bootstrap estimators, which are arguably more robust to such noise, lead to similar findings.

[12] In Online Appendix H.1, we show that this finding is robust when we separately estimate the conditional choice model for each job category. One might argue that all coefficients will be significant due to the large size of our sample. However, we observe that the estimated coefficients for many variables (profile completed, some freelancers' regions, and beauty) are not statistically different from zero.

[13] The primary goal of our observational study is descriptive (i.e., exploring the role of profile pictures in hiring decisions). As such, we believe that comparing models using in-sample fit measures (such as AIC and BIC) rather than out-of-sample prediction accuracy would be sufficient in our setting (Boughanmi et al. 2016). One might still wonder whether perceived job fit can help improve out-of-sample prediction accuracy. We carried out such an exercise in Online Appendix H.2.

[14] Arguably, employers may form two additional relevant perceptions from freelancers' profile pictures: perceived competence and perceived professionalism. We created these labels following a similar procedure as in Section 3.2, and included them in the choice models. As we show in Online Appendix H.3, the inclusion of these two additional controls does not change the main results presented in this section.

[15] The nonsignificant correlation between the number of reviews and perceived job fit does not contradict our main finding that freelancers who "look the part" are more likely to be hired. The total number of reviews is likely to be explained by the number of years the freelancers have worked on the platform. Indeed, perceived job fit is positively and significantly correlated with the change in the log number of reviews freelancers experienced during our sample period ($r = 0.018$, $p < 0.01$), which reinforces the findings that perceived job fit is associated with better hiring outcomes.

[16] This attribute can also be useful to replace the application description used in the platform (Figure 1) because it is challenging to manipulate preset levels of application descriptions with multiple versions of texts. In contrast, certification is a clean manipulation that summarizes the key information the freelancer can include in the text, such as "I have experience with WordPress." Moreover, the results from the observational study indicate that certifications correlate significantly with hiring outcomes.

[17] U.S. Equal Employment Opportunity Commission. Source: https://www.eeoc.gov/prohibited-employment-policiespractices#pre-employment_inquiries.

References

Adomavicius G, Bockstedt J, Curley S, Zhang J (2011) Recommender systems, consumer preferences, and anchoring effects. *CEUR Workshop Proc.*, vol. 811, 35–42.

Aribarg A, Burson KA, Larrick RP (2017) Tipping the scale: The role of discriminability in conjoint analysis. *J. Marketing Res.* 54(2): 279–292.

Athey S, Karlan D, Palikot E, Yuan Y (2022) Smiles in profiles: Improving fairness and efficiency using estimates of user preferences in online marketplaces. Working Paper No. w30633, National Bureau of Economic Research.

Bairathi M, Zhang X, Lambrecht A (2022) The value of platform endorsement. Preprint, submitted June 23, https://dx.doi.org/10.2139/ssrn.4144605.

Benlian A, Titah R, Hess T (2012) Differential effects of provider recommendations and consumer reviews in e-commerce transactions: An experimental study. *J. Management Inform. Systems* 29(1):237–272.

Benson A, Sojourner A, Umyarov A (2020) Can reputation discipline the gig economy? Experimental evidence from an online labor market. *Management Sci.* 66(5):1802–1825.

Boughanmi K, Kohli R, Jedidi K (2016) Hit or fit: The tradeoff between maximizing likelihood and hit rate. Preprint, submitted August 3, http://dx.doi.org/10.2139/ssrn.4161908.

Burgoon JK (1978) Attributes of the newscaster's voice as predictors of his credibility. *Journalism Mass Comm. Quart.* 55(2):276–300.

Chan J, Wang J (2018) Hiring preferences in online labor markets: Evidence of a female hiring bias. *Management Sci.* 64(7): 2973–2994.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proc. 22nd ACM Sigkdd Internat. Conf. Knowledge Discovery Data Mining*, 785–794.

Davis DR, Dingel JI, Monras J, Morales E (2019) How segregated is urban consumption? *J. Political Econom.* 127(4):1684–1738.

Doleac JL, Stein LCD (2013) The visible hand: Race and online market outcomes. *Econom. J. (London)* 123(572):F469–F492.

Edelman BG, Luca M (2014) Digital discrimination: The case of airbnb. com. Working paper, Harvard Business School.

Ert E, Fleischer A, Magen N (2016) Trust and reputation in the sharing economy: The role of personal photos in airbnb. *Tourism Management* 55:62–73.

Fagerstrøm A, Pawar S, Sigurdsson V, Foxall GR, Yani-de Soriano M (2017) That personal profile image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behavior on airbnb™. *Comput. Human Behav.* 72:123–131.

Feldman JM, Lynch JG (1988) Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *J. Appl. Psych.* 73(3):421.

Filippas A, Horton JJ, Golden JM (2022) Reputation inflation. *Marketing Sci.* 41(4):733–745.

Gai PJ, Klesse A-K (2019) Making recommendations more effective through framings: Impacts of user-vs. item-based framings on recommendation click-throughs. *J. Marketing* 83(6): 61–75.

Hannák A, Wagner C, Garcia D, Mislove A, Strohmaier M, Wilson C (2017) Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. *Proc. ACM Conf. on Comput. Supported cCooperative Work and Social Comput*. 1914–1933.

Hartmann J, Heitmann M, Schamp C, Netzer O (2021) The power of brand selfies. *J. Marketing Res.* 58(6):1159–1177.

Kawaguchi K, Uetake K, Watanabe Y (2019) Effectiveness of product recommendations under time and crowd pressures. *Marketing Sci.* 38(2):253–273.

Kokkodis M, Ipeirotis PG (2016) Reputation transferability in online labor markets. *Management Sci.* 62(6):1687–1706.

Kokkodis M, Papadimitriou P, Ipeirotis PG (2015) Hiring behavior models for online labor markets. *Proc. 8th ACM Internat. Conf. on Web Search and Data Mining*, 223–232.

Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. *Comm. ACM* 60(6):84–90.

Krumhuber E, Manstead ASR, Cosker D, Marshall D, Rosin PL (2009) Effects of dynamic attributes of smiles in human and

Downloaded from informs.org by [207.151.52.80] on 09 January 2023, at 10:29 . For personal use only, all rights reserved.

synthetic faces: A simulated job interview setting. *J. Nonverbal Behav.* 33(1):1–15.

Lennan C, Nguyen H, Tran D (2018) Image quality assessment. Accessed May 10, 2021, https://github.com/idealo/image-quality-assessment.

Leung W, Zhang Z, Jibuti D, Zhao J, Klein M, Pierce C, Robert L, Zhu H (2020) Race, gender and beauty: The effect of information provision on online hiring biases. *Proc. CHI Conf. on Human Factors in Comput. Systems*, 1–11.

Liang C, Shi Z, Raghu TS (2019) The spillover of spotlight: Platform recommendation in the mobile app market. *Inform. Systems Res.* 30(4):1296–1318.

Liu X, Lee D, Srinivasan K (2019) Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *J. Marketing Res.* 56(6):918–943.

Ludwig S, Herhausen D, Grewal D, Bove L, Benoit S, De Ruyter K, Urwin P (2022) Communication in the gig economy: Buying and selling in online freelance marketplaces. *J. Marketing* 86(4):141–161.

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, et al. (2020) From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence* 2(1):56–67.

Luo L, Kannan PK, Ratchford BT (2008) Incorporating subjective characteristics in product design and evaluations. *J. Marketing Res.* 45(2):182–194.

Lynch JG Jr, Marmorstein H, Weigold MF (1988) Choices from sets including remembered brands: Use of recalled attributes and prior overall evaluations. *J. Consumer Res.* 15(2):169–184.

McFadden D (1973) Conditional logit analysis of qualitative choice behavior. Zarembka P, ed. *Frontiers in Econometrics* (Academic Press, New York), 105–142.

Olivola CY, Todorov A (2010a) Elected in 100 milliseconds: Appearance-based trait inferences and voting. *J. Nonverbal Behav.* 34(2):83–110.

Olivola CY, Todorov A (2010b) Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *J. Experiment. Soc. Psych.* 46(2):315–324.

Olivola CY, Sussman AB, Tsetsos K, Kang OE, Todorov A (2012) Republicans prefer republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. *Soc. Psych. Personality Sci.* 3(5):605–613.

Orme B (2002) Formulating attributes and levels in conjoint analysis. Sawtooth Software research paper series, Sawtooth Software, Sequim, WA.

Pope DG, Sydnor JR (2011) What's in a picture? Evidence of discrimination from prosper. com. *J. Human Resources* 46(1): 53–92.

Rao VR (2014) *Applied Conjoint Analysis* (Springer, New York).

Senecal S, Nantel J (2004) The influence of online product recommendations on consumers' online choices. *J. Retailing* 80(2):159–169.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A simple way to prevent neural networks from overfitting. *J. Machine Learn. Res.* 15(1):1929–1958.

Sun M (2012) How does the variance of product ratings matter? *Management Sci.* 58(4):696–707.

Watson J, Ghosh AP, Trusov M (2018) Swayed by the numbers: The consequences of displaying product review attributes. *J. Marketing* 82(6):109–131.

Wei X, Stillwell D (2017) How smart does your profile image look? Estimating intelligence from social network profile images. *Proc. 10th ACM Internat. Conf. on Web Search and Data Mining*, 33–40.

Yoganarasimhan H (2013) The value of reputation in an online freelance marketplace. *Marketing Sci.* 32(6):860–891.

Zhang M, Luo L (2022) Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from yelp. *Management Sci.*, ePub ahead of print April 4, https://doi.org/10.1287/mnsc.2022.4359.

Zhang X, Zhao J, LeCun Y (2015) Character-level convolutional networks for text classification. *Adv. Neural Inform. Processing Systems*, 649–657.

Zhang S, Lee D, Singh PV, Srinivasan K (2021) What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Sci.* 68(8):5644–5666.

Zhang S, Friedman E, Zhang X, Srinivasan K, Dhar R (2020) Serving with a smile on airbnb: Analyzing the economic returns and behavioral underpinnings of the host's smile. Preprint, submitted September 14, https://dx.doi.org/10.2139/ssrn.3692623.