RESEARCH PAPER

The role of contextual and contentual signals for online trust: Evidence from a crowd work experiment

Rense Corten[1] · Judith Kas[2] · Timm Teubner[3] · Martijn Arets[2]

Abstract

Platform workers can typically not take their ratings from one platform to another. This creates lock-in as building up reputation anew can come at prohibitively high cost. A system of portable reputation may mitigate this problem but poses several new challenges and questions. This study reports the results of an online experiment among 180 actual clients of five gig economy platforms to disentangle the importance of two dimensions of worker reputation: (1) contextual fit (i.e., the ratings' origin from the same or another platform) and (2) contentual fit (i.e., the ratings' origin from the same or a different job type). By and large, previous work has demonstrated the potential of imported ratings for trust-building but usually confounded these two dimensions. Our results provide a more nuanced picture and suggest that there exist two important boundary conditions for reputation portability: While imported ratings can have an effect on trust, they may only do so for matching job types and in the absence of within-platform ratings.

Keywords Gig economy · Crowd work · Platform economy · Reputation portability · Trust · Signaling theory · Lock-in effects

Introduction

In the platform economy, reputation is crucial. A large number of studies show that individuals with good ratings and reviews are trusted more (Boero et al., 2009; Bolton et al., 2004; Charness et al., 2011; Duffy et al., 2013; Frey & Van

Responsible Editor: Jörn Altmann

✉ Timm Teubner
teubner@tu-berlin.de

Rense Corten
R.corten@uu.nl

Judith Kas
judithkas@outlook.com

Martijn Arets
mgcarets@hhs.nl

[1] Department of Sociology, Utrecht University, Padualaan 14, 3584, CH, Utrecht, the Netherlands

[2] Utrecht, Netherlands

[3] Trust in Digital Services, Einstein Center Digital Future, TU Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany

De Rijt, 2016). Importantly, ratings are usually collected, displayed, and confined *within* a single platform. Uber drivers cannot simply take their reputation to Lyft if they want to multi-home or switch. Getting the first review on a new platform is difficult and costly, as clients have little reason to trust workers who have not received any ratings and/or reviews yet (Przepiorka, 2013). This potentially creates a lock-in: workers cannot easily switch platforms as the costs of building a new reputation from scratch can be prohibitively high (Dellarocas, 2010; Demary, 2015). However, such lock-ins impair innovation and may lead to excessive rents for incumbents (e.g., Farell & Klemperer, 2007; Katz & Shapiro, 1994) which is considered problematic. In this sense, the German government has demanded that "platform workers must be able to take their reviews to another platform" (Lambrecht & Heil, 2020). Other voices have joined in: In March 2021, the German Trade Union Confederation published a position paper on platform work, demanding reputation portability as a mechanism for workers to circumvent potential platform lock-in (DGB, 2021). Also, the European Commission (EU, 2017) has early on suggested to study the "mechanisms for reputation portability, assessing its advantages and disadvantages and technical, legal and practical feasibility" (p. 131). Also, the International

Labour Organization (ILO) and the Dutch Social and Economic Council recommend reputation portability as a solution against platform lock-in (Choudary, 2018; Rani et al., 2021; Sociaal-Economische Raad, 2020).

Arguably, the problem of potential lock-in is particularly relevant to crowd work, given that platform workers may depend on their employability for their livelihood (Schor et al., 2020). Along with the number of platform workers in Europe, it is likely that the dependence on platforms to earn a living is also increasing (Urzi Brancati et al., 2020). A system that would allow workers to take their ratings from one platform to another may hence mitigate this problem. And indeed, several initiatives and infrastructures for such portable reputation have emerged in recent years (Hesse et al., 2022; Hesse & Teubner, 2020a; Teubner et al., 2020).

It is important to note that the usefulness and applicability of imported and within-platform ratings may depend on several factors. First, different platforms may facilitate *different types of jobs*. For instance, on platforms such as Taskrabbit, job categories include furniture assembly, home cleaning, delivery services, yard work, and many more. Ratings from one job type may not be as useful in another category, because different jobs require different sets of skills and traits (Kokkodis & Ipeirotis, 2016). Clients may thus learn little about a worker's future behavior and performance if there is little overlap between the demonstrated and required skills and traits between jobs.

Second, imported ratings inherently root from a *different platform* — even though they may be associated with an identical or similar job. This may affect their usefulness in several ways. For instance, there may be differences and uncertainties about the reputation system and the way in which ratings are constructed on the other platform. This limits clients' understanding of the norms, credibility, and relevance of such imported ratings and makes it difficult to assess their face value. Even if both platforms formally use the same scale (e.g., 1–5 stars), the implicit rules, assumptions, and standards may be quite different across contexts. For example, Uber clearly indicates that a five-star review is the default, while Airbnb does not provide specific information and lower rating scores are more common (although still rare). Naturally, both effects may coincide, for instance, when a worker aims to transfer their ratings from Uber (i.e., matching taxi drivers to riders) to CharlyCares (matching babysitters to families).

While providing important first insights, previous studies on portable reputation systems have not distinguished between ratings that originate from a different platform and ratings about a different type of job. Instead, they have looked at different platforms that facilitate matching of workers to clients for different jobs — hence confounding these two potential effects. For example, in both Teubner et al. (2020) and Otto et al. (2018), respondents were asked whether or not they would engage with an Airbnb host or BlaBlaCar driver who either had no ratings or an imported rating from some other platform. Hence, the treatment conditions confounded the reputation's source as well as job type it referred to. Given that — unlike on Airbnb or Uber — gig economy platforms usually offer many different job types, we argue that this aspect needs further disentanglement and exploration.

In a broader theoretical perspective, our study contributes to the understanding of the *robustness* of reputation effects under conditions of imperfect information (e.g., Bolton et al., 2004, 2005; Granovetter, 1985). The theory on mechanisms of reputation formation and its effects on trust is well-developed, drawing on game-theoretical argumentation and signaling theory in particular (Tadelis, 2016). Empirically, these arguments have been studied extensively for conditions under which reputational information is readily and reliably available in well-defined contexts (Bolton et al., 2004; Diekmann et al., 2014; Jiao et al., 2021; Resnick & Zeckhauser, 2002, 2006), but less is known about the robustness of reputation effects against various sources of "noise" that render the relationship between reputational signals and the underlying behavior less than perfect. The transfer of reputation scores across platforms and across job types constitutes two such sources of imperfection. Existing work on reputation transferability, which relies mostly on signaling theory (Hesse et al., 2022; Otto et al., 2018; Teubner et al., 2020), does not yet provide theoretical insights in the role of these two dimensions of imperfect information. While the core theory argues that that signaling serves as a means for individuals to convey information about their attributes, traits, or intentions to others, we expand this conceptualization by two factors relevant to the online crowd work business: *contextual* and *contentual fit*. We argue that these boundary conditions should be taken into account when considering reputational signals and why/when these are — or are not — effective for creating trust (in the respective signal's sender).

This paper's main contribution is hence the theoretical conceptualization of contextual and contentual fit as relevant dimensions of imperfect information as well as the empirical disentanglement of these two dimensions. This improves our understanding of the boundary conditions under which imported reputation can be effective. To do so, we conducted an online experiment in which we presented platform clients with the profiles of gig workers. We systematically varied (a) the presence of imported and within-platform ratings and (b) the type of job that the ratings have been acquired in — allowing us to systematically test and distinguish the effects of each. In contrast to earlier research based on samples of general Internet users (; Hesse et al., 2022; Otto et al., 2018; Teubner et al., 2020) or observational studies (Norbutas et al., 2020b), we experimentally study the behavior of *actual platform clients*, which yields higher external validity than earlier work. We collaborated with five gig economy platforms to get access to a sample of such clients (from *CharlyCares, Helpling, Level.works, Temper, Young-Ones*). With this study, we hence aim to answer the questions to

what extent ratings from a different job and/or a different platform can serve as a signal of trustworthiness and, based on this, how it can help alleviate workers' lock-in to a certain platform.

Context and theory

The platform economy and trust

The "platform economy" has emerged as an umbrella term that encompasses several activities, including selling, exchanging, borrowing, and renting of goods and services. In this paper, we specifically focus on crowd work (or gig) platforms, which mediate paid tasks carried out by independent contractors (or workers) (Koutsimpogiorgos et al., 2020). On such platforms, clients (either organizations or individuals) typically post job offers for which individual workers can apply for. Clients can then choose one or more of the workers to carry out the job, unlike on on-demand platforms such as Uber, where an algorithm assigns workers to clients. Clients and workers usually do not meet in person before the start of the job, and they often interact only once (Koutsimpogiorgos et al., 2020). Thus, there is substantial information asymmetry between workers and clients which causes a distinct *trust problem*: as information asymmetry allows opportunistic behavior by workers, clients have reason to refrain from engaging in transactions, which is a suboptimal outcome for both (cf. Akerlof, 1970). The pervasiveness of this trust problem in the platform economy is seen as a key challenge, and has given rise to a large trust, reputation, and signaling literature (Cook et al., 2009; Dellarocas, 2003; Kollock, 1999; Resnick & Zeckhauser, 2002; Tadelis, 2016; ter Huurne et al., 2017). Arguably, the trust problem is particularly relevant in situation with high degree of economic, social, and/or physical exposure (e.g., in-house cleaning or babysitting), and for jobs that require high levels of skill or are more difficult to monitor. For the scope of this study, we define trust as (a client's) willingness to accept such vulnerability due to others' (i.e., a worker's) actions based on expectations about their intentions and skills (Mayer et al., 1995; Rousseau et al., 1998). In order for a client to trust a worker, they need to believe that the worker is capable of properly carrying out the job, and also willing to do so in a satisfactory way. Without any information about a worker, a client simply cannot know in advance which worker will and which one will not do a good job.

Reputation systems

This is where reputation systems come in. To overcome the above stated problem, most platforms make use of reputation systems that collect, aggregate, and distribute feedback about individuals' past behavior (Lehdonvirta, 2022; Resnick et al., 2000; Resnick & Zeckhauser, 2006; Tadelis, 2016). One objective of reputation system design is to give clients the opportunity to distinguish between trustworthy and untrustworthy workers, even when these workers are strangers with whom clients have never worked before (Nannestad, 2008; Weigelt & Camerer, 1988). In particular, many platforms allow clients to provide a rating on a scale from one to five stars and to write a text review after a transaction is completed.

The effectiveness of such systems in fostering trust is often understood in terms of signaling theory (e.g., Akerlof, 1970; Otto et al., 2018; Przepiorka, 2013; Tadelis, 2016) where workers' (online) reputation is considered to credibly signal information about them. A trustworthy worker can be defined as successfully completing a given job because they are competent (i.e., hold the required skills and knowledge), benevolent (i.e., hold a concern for the client's interests), and have integrity (i.e., they adhere to a set of sound principles and keep their word). A positive (average) rating signals that a worker successfully completed one or more jobs in the past, while a negative rating signals that there have occurred problems. Assuming workers' past behavior to be related to their future behavior, reputation thus informs clients about the likelihood that a worker will successfully complete a job.

A large number of studies have shown that individuals with a good reputation are trusted more (Boero et al., 2009; Bolton et al., 2004; Charness et al., 2011; Diekmann et al., 2014; Duffy et al., 2013; Edelman & Luca, 2014; Frey & Van De Rijt, 2016; Przepiorka, 2013; Teubner et al., 2017). If workers with better ratings are trusted more and if a worker wishes to participate in more transactions in the future, they will be motivated to keep up a good reputation by acting trustworthy (Benard, 2013; Buskens & Raub, 2002; Buskens & Weesie, 2000; Charness et al., 2011; Cheshire, 2007; Kroher & Wolbring, 2015; Rooks et al., 2006). While reputation systems have been criticized, for instance, in view of aspects such as reputation inflation, staticity, or issues of attribution (Kokkodis, 2021), the positive effect of good ratings on trust is well-established in the literature (Cui et al., 2020; Jiao et al., 2021; Kas et al., 2022; Tadelis, 2016; ter Huurne et al., 2018; Teubner et al., 2017; Tjaden et al., 2018).

However, an inherent feature of using reputation systems is that they make it difficult for newcomers to enter the market, as these do not have a track record to show for. Getting a first review is of utmost importance to participate in future interactions, because ratings tend to cascade (Duffy et al., 2013; Frey & Van De Rijt, 2016). Individuals who already have one or more positive reviews are more likely to be selected for upcoming jobs, meaning that there will occur *Matthew Effect* dynamics (Merton, 1968; van de Rijt et al., 2014). Newcomers on a platform need to convince clients that they are trustworthy without relying on ratings and reviews. One commonly observed way to acquire a first

Table 1 Related literature on reputation portability

Study	Focus	Method	Context	Sample
Hesse et al. (2022)	Interplay of onsite and imported ratings of different valence	Experimental online survey	Accommodation sharing, Airbnb	493; Prolific
Ciotti et al. (2021)	Crowd workers' switching behavior between online labor markets	Experiment	Crowd work	1622; AMT
Teubner et al. (2020)	Source-target (platform) fit as a driver of effectiveness of imported reputation	Experimental online survey	eBay, Uber, Airbnb, BlaBlaCar	403; Prolific
Hesse et al., (2020)	Rating discrepancy as an inhibiting factor of trust-building	Experimental online survey	Accommodation sharing, Airbnb	305; Prolific
Hesse and Teubner (2020b)	Effect of imported rating on sales	Observational	E-commerce (Bonanza.com)	4506; Bonanza
Hesse and Teubner (2020a)	Conceptual framework on the involved parties and relations in reputation portability	Conceptual	—	—
Norbutas et al. (2020b)	Sellers' migration behavior between crypto market platforms	Observational	Dark Web markets: Abraxas, Agora, AlphaBay	7593; cryptomarket vendors
Teubner et al. (2019)	Conceptual differentiation of trust, trust transfer, and reputation transfers	Conceptual/Survey	28 online platforms	494; Prolific
Otto et al. (2018)	Effect of importing reputation from different platform	Experimental online survey	Airbnb, BlaBlaCar	141; students
Zloteanu et al. (2018)	Effect of showing imported reputation from different platforms	Experimental online surveys	Accommodation sharing, eBay, Uber	120, 117, 189; AMT
Kokkodis and Ipeirotis (2016)	Estimation of crowd worker quality based on past job evaluations from different categories	Observational	Crowd work (oDesk)	>1 million transactions; oDesk

rating is by offering lower prices/wages (Przepiorka, 2013), but on many gig platforms, this is not possible as prices are set by clients (e.g., Amazon Mechanical Turk) or the platform itself (e.g., Uber). As a consequence, it is difficult for workers to switch to a new platform — or to work add an additional platform to their work portfolio (i.e., multi-homing) (Dellarocas, 2010; Demary, 2015).

Portable reputation

Portable reputation systems may be a solution to this issue but empirical research on this matter has emerged only recently (see Table 1). While there is some evidence that ratings may increase trust across contexts, for example, across different experimental games (Fehrler & Przepiorka, 2013) or different tasks (Kokkodis & Ipeirotis, 2016), research that looks more specifically at cross-platform reputation is still scarce (Teubner et al., 2019). For instance, Zloteanu et al. (2018) found that survey respondents seem to value within-platform over cross-platform information, but did not study the impact of cross-platform reputation on trust directly. More direct evidence for the effectiveness of imported ratings comes from online experiments, in which respondents assess

fictional provider profiles. Otto et al. (2018) tested the general hypothesis derived from signaling theory that imported ratings increase trust, and indeed found a positive effect. More recent online experiments confirm this general notion also when both within-platform and imported ratings are present (Hesse et al., 2022) but also identified limiting conditions such the similarity of source and target platform (Teubner et al., 2020) and the *value* of imported ratings; in particular, the effect of imported ratings may be detrimental when their scores are relatively low (Hesse et al., 2022). Moreover, rating discrepancy (i.e., differences between onsite and imported ratings), regardless into which direction, seem to have detrimental effects on trust compared to more similar rating scores with same average value (Hesse et al., 2020).

Empirical research that studies the effectiveness of imported ratings in field settings is even scarcer. Norbutas et al. (2020b) study the portability of reputation in the context of Dark Web drug markets, and find that sellers who migrate between different markets indeed benefit from importing ratings from previous interactions on other markets. In contrast, Hesse and Teubner (2020b) do not find clear effects of imported ratings on the e-commerce platform Bonanza.com.

Finally, we note that earlier research on reputation portability has focused on the sharing economy (e.g., Airbnb, BlaBlaCar) and e-commerce platforms (e.g., eBay, Bonanza, Dark Web markets), but much less so on the gig economy or crowd work. An exception is Ciotti et al. (2021) who, however, study the behavior of platform *workers* rather than *clients*, as is the focus of our study.

Over the years, a number of bottom-up initiatives have emerged that explore technical and entrepreneurial possibilities of creating independent reputation profiles that would help individuals to take ratings from one platform to the other, with varying levels of success (Teubner et al., 2020). These solutions may allow workers to directly integrate rating scores from different platforms in their worker profiles. Just like within-platform ratings, imported ratings can serve as a signal of past behavior that clients can use to distinguish between trustworthy and untrustworthy workers. Such portable reputation would thus allow workers to enter a new platform with a head start: they do not need to completely start all over as their imported ratings signal that they have been trustworthy in the past.

Current theory on reputation mechanisms, however, provides an ambiguous answer as to the expected effectiveness of such imported reputation for fostering trust. It depends on the specific assumptions made regarding the information available to the actors and, consequently, on the impact of factors that render this information less than perfect. On the one hand, "classic" game-theoretical models of reputation (Kandori, 1992) that rely on the existence of subgame-perfect Nash equilibria require that a complete history of all previous actions of the relevant parties is available to the actors involved (i.e., "perfect information" in game-theoretical terms). In such models, the effectiveness of reputation mechanisms is highly sensitive to imperfections in the availability of information. For instance, in reputation systems, information on the complete "history of play" is typically not available, leading to the prediction that reputation systems should be less effective in fostering trust than direct observation (Bolton et al., 2005). Extending this argument to imported reputation, where information can be expected to be even less perfect, suggests that imported reputation should be less effective in fostering trust than within-platform reputation.

Competing models (e.g., "action discrimination" models such as the image scoring model (Nowak & Sigmund, 1998)) do not rely on subgame perfection and propose that the availability of information about a partner's most recent actions is sufficient to maintain cooperation. Such models thus predict that reputation mechanisms are more robust to imperfect information. They are therefore more optimistic about the effectiveness of reputation systems (Bolton et al., 2005) or other imperfect sources of reputation such as gossip (Sommerfeld et al., 2007). Based on this line of reasoning, one may suggest that imported reputation can be as effective as within-platform reputation. Experimental evidence indicates that, while models relying on subgame perfection are too strict in their information requirements, action discrimination models are too lenient (Bolton et al., 2005).

Similarly, the existing theoretical literature contains contradictory views on the impact of the exact *source* of information. On the one hand, game-theoretical models typically assume that "all information is equal"; that is, it is the *content* of information that matters, not its source. This implies that reputational information imported from other platforms should be as effective as within-platform reputation, on the condition that it is equally accurate. Sociological theories of reputation, on the other hand, emphasize that the source of information does in fact matter. For example, Granovetter (1985) argues that first-hand information should be expected to have a stronger impact than information transferred by third parties, as among other reasons, it is deemed to be more reliable (Bolton et al., 2004; Norbutas et al., 2020a). As reputational information imported from different platforms increases the social distance between the source of the information and its receiver even further, this would suggest that imported reputation is less effective than within-platform reputation.

At the individual level, the imperfection of information introduced by importing reputational information from different platforms can be understood in terms of the psychological concept of *categorization* (Shaw, 1990; Teubner et al., 2020). This describes a heuristic in which individuals, faced with large amounts of information, tend to process novel information that is similar to earlier information in similar ways. Thus, the more familiar the receiver is with the source of information, the more likely this information is to be effective. Conversely, if the source is less familiar (in our case, because the source is a different platform), it is more likely to be rejected. Since the large majority of ratings on online platforms are positive ratings (Schoenmueller et al., 2018) and because it is unlikely that workers would bring negative ratings to a new platform, we focus on positive ratings in this study. The baseline expectation, given the extensive literature on reputation effects in online markets, is that within-platform ratings positively impact trust. Assuming that this logic can be extended to imported reputation, we formulate the following (to-be-verified) baseline assumptions:

H1a: The presence of a within-platform rating is positively associated with clients' trust in workers.
H1b: The presence of an imported rating is positively associated with clients' trust in workers.

There are, however, a number of factors that may limit the relative effectiveness of imported ratings. Given the diversity

of theoretical views in the literature, we argue that the framework of signaling should take into account the boundary conditions of *contextual* and *contentual fit*. First, using reputational information from the same platform environment establishes *contextual fit* for the recipient. Within this (same) context, the recipient will — broadly speaking — be more familiar with the type of signal, its characteristics, what different scores mean, etc. — and is hence likely to put greater weight on the signal. In contrast, cross-contextual signals introduce an additional layer of noise that is likely to impair the recipient's capability (and hence, their willingness) to fully appreciate the signal. It is important to note here that the signal itself does not become noisier per se, but that — given the circumstances of its interpretation (i.e., in- or out-of-context), noise stems from the interpretation process.

Second, using reputational information from the same job or closely related jobs from the same category provides *contentual fit* for the recipient (i.e., the client on the crowd work platform). Observing such a signal speaks to its applicability for the task at hand as there will be substantial overlap with regard to the requirements, subject matter, and contents of the job. In contrast, non-contentual signals will introduce an additional layer of noise and uncertainty with regard to how well the signaled characteristics will transfer to the targeted job. Hence, different combinations of source and target jobs will allow for different levels of transferability. While some jobs share more characteristics (e.g., Uber driver and delivery driver), other combinations will barely have any commonalities beyond rudimentary features such as reliability, punctuality, etc. (e.g., Uber driver vs. baby sitter vs. gardening).

Contextual fit

By definition, imported ratings have been earned on a different platform. Clients who are evaluating another person's profile may not know the source platform(s), and may therefore not be able to assess the usefulness of the imported rating(s). One example of differences between platforms are the mechanisms and implicit norms in how clients and workers rate each other. Some platforms, such as Uber, clearly indicate that a five-star review is the default, and only when something went wrong, the passenger should give a lower rating. This results in a very skewed distribution of ratings (Liu et al., 2021). On other platforms, such as Airbnb, no clear anchor is provided. Despite these differences, on many platforms, the distribution of ratings is highly skewed toward mostly 4- and 5-star ratings (Schoenmueller et al., 2018).

Another difference is whether or not platforms fully describe the meaning of every point on the rating scale. For example, on Airbnb, a description of the meaning of different numbers of stars is provided (1 star means "terrible," 3 stars means "ok," 5 stars means "amazing"). Other platforms only provide the scale without further explanation. Further

differences in reputation systems include, for instance, the scale used, simultaneous vs. sequential rating schemes, rating withdrawal rules (Bolton et al., 2023), whether or not the entire distribution is conserved or only an aggregate score is provided, whether the source platform operates an "open" reputation system where basically anyone can leave a review for a certain provider or product (e.g., Yelp, Google, Amazon), or whether the system is "closed" (such as on most sharing- and crowd work platforms). These differences may result in different rating distributions on different platforms, and therefore in different interpretations of ratings (Schoenmueller et al., 2018). When clients do not know how a rating is constructed on the source platform, they may find it less informative and, consequently, attach less weight to it than to (more familiar) within-platform ratings. Based on this, we hypothesize that:

> H1c: A within-platform rating has a stronger effect on clients' trust in the worker than an imported rating.

Contentual fit

A key objective of our study is to disentangle the effects of different causes of noise in the reputation signals contained in imported ratings, in particular, on the one hand, differences between origin- and target platform, and on the other hand, differences in the jobs on which reputation is earned. To be able to evaluate these effects separately, we develop a separate hypothesis on the effects of ratings earned for similar jobs, versus ratings earned for different jobs. Platforms differ to the extent that they specialize in certain types of jobs. While some platforms specialize in only one type of work (e.g., Uber for drivers, Helpling for cleaners), other platforms allow workers to perform multiple types of jobs (e.g., Taskrabbit, Temper). Similar tasks require similar skills. Hence, information about a worker's past performance on a similar job may help clients to learn whether the worker has the required skills, and whether he/she has proven to be trustworthy in the past. But even when the job for which a client intends to hire a worker is different from the job in which the worker earned a rating, the rating may still be informative to the client in terms of more general qualities (Kokkodis & Ipeirotis, 2016). Even though jobs may be different, they may share requirements with regard to some skills, benevolence, and integrity. For larger overlaps of requirements, information from the past will be more informative. If someone tries to find a job as a baby sitter via a platform, a rating earned with another baby sitting job will certainly be of more use than a rating earned in delivery work. Thus, while we expect that both same-job ratings and different-job ratings have positive effects on trust, we hypothesize that this effect is stronger for same-job ratings than for other-job ratings (Fig. 1):

Fig. 1 Overview of hypotheses

H2a: The presence of a same-job rating is positively associated with clients' trust in workers.

H2b: The presence of an other-job rating is positively associated with clients' trust in workers.

H2c: Clients trust workers who have same-job ratings more than they trust workers with other-job ratings.

Interaction effects

When worker use multiple platforms at the same time, they may leverage both within-platform ratings and imported ratings. As outlined in hypotheses H1c und H2c, we posit that "closer" and more specific information (i.e., within-platform and same-job) will have stronger effects on trust than more distant and less specific information (i.e., imported and other-job). Moreover, we expect that the presence of the first (within-platform, same-job) will be dominant and hence reduce the impact of the second (imported, other-job). In other words, when clients only have access to imported ratings, and/or other-job ratings, their best strategy would be to rely on that information. However, when better information is available, there will be less need to rely on this "inferior" information. We therefore hypothesize that:

H1d: The effect of an imported rating on clients' trust is stronger in the absence of a within-platform rating.

H2d: The effect of an other-job rating on clients' trust is stronger in the absence of a same-job rating.

Method

To evaluate our hypotheses, we conducted an online experiment among clients of five gig economy platforms in the Netherlands: Helpling, Temper, YoungOnes, CharlyCares, and Level.works. All respondents had previously hired workers via one of the platforms and provided consent before starting the experiment. We asked participants for which job they last hired a worker, and then, we asked them to imagine that they would seek to hire someone for that same job again. We then showed them the profile of a hypothetical worker.

Treatment design

For the features of this worker profile, we used a 3-by-3 treatment design. In the first dimension, we varied the availability and source of *within-platform rating* (none, same job, other job). Similarly, in the second dimension, we varied the availability and source of an *imported rating* (none, same job, other job). This yielded a total of $3 \times 3 = 9$ treatment conditions (Fig. 2). All of these combinations were evaluated (full-factorial design). Moreover, each participant was assigned to only one single treatment condition (between-subjects design). Note that our main variables of interest (e.g., whether there is a within-platform rating, an imported rating, a same-job rating, and/or an other-job rating) as well as their overlaps cover different subsets of this main treatment matrix.

Stimulus material

We started the survey by asking respondents about the last job they hired a worker for. Based on that and the platform they were using, we then showed them the profile of the hypothetical worker with reputation according to one out of the nine treatment conditions (see Fig. 3 for an example). Each participant was randomly assigned one of four (blurred) worker profile images, two of which were female, and two were male.

Procedure

We collaborated with the five aforementioned platforms (Temper, YoungOnes and Level.works, Helpling, CharlyCares), which sent emails to their clients that had previously hired a worker via their platform. A total of 9663 emails were sent between June and October 2021. Between 1 and 2 weeks after the first invitation, a reminder was sent. Overall, this yielded 197 responses (response rate: 2%). As an attention check, we asked participants to click the button in the center of the Likert scale. Ten participants were excluded because they did not pass this check. Seven more participants were excluded because they encountered a

Fig. 2 Treatment matrix (including observation numbers and main variables' treatment partitioning).Because the platforms Helpling and CharlyCares each focus on a single type of job, it was not possible to assign respondents from those platforms to the within-platform other-job conditions, resulting in lower observation numbers in these cells. "Within" and "Imported" denote the rating's source platform, while "none" vs. "same" and "other" refer to the job-match condition
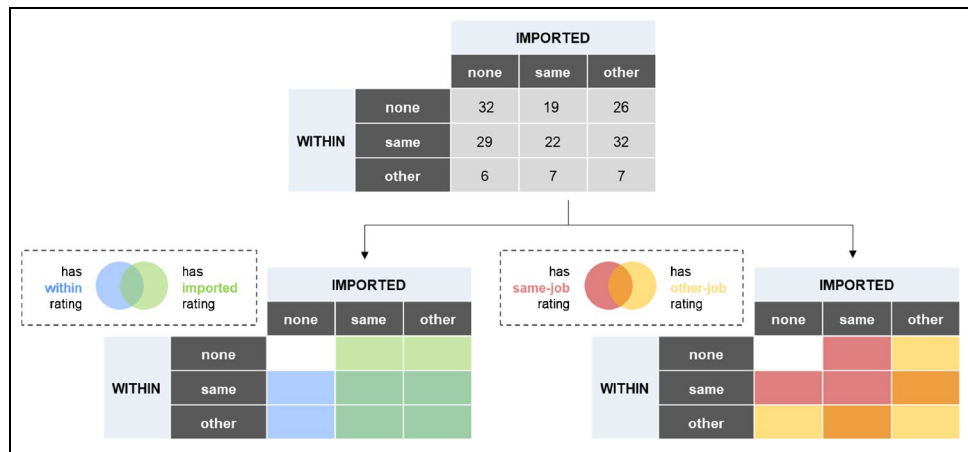
programming error in the experiment. The final sample size was hence $n = 180$. Among all participants that completed the study, we organized a lottery of 25 vouchers of 20 EUR each and an iPad. Participants could complete the survey either in English or in Dutch. Before commencing with the experiment, we obtain ethical approval by the ethics board at the principal investigator's institution (Utrecht University).

Measurement instrument

We asked participants how trustworthy they deemed the shown worker profile (*trusting belief*) and how likely they would be to hire them (*trusting intention*). These (single-item) questions were asked on 7-point Likert scales. Moreover, we surveyed participants on a series of demographic and other control variables (e.g., disposition to trust, familiarity with crowd work, perceived realism of the experiment). The full survey is provided in

Appendix 1. We generated two sets of binary variables that indicated which type of ratings participants saw in their respective treatment condition. Two binary variables indicate whether the respondent saw within-platform ratings and/or imported ratings. The second set of variables indicated whether participants saw same-job ratings and/ or other-job ratings. We categorized all jobs in six categories: babysitting, hospitality, cleaning, delivery/logistics, promotion, and retail. The full list of jobs can be found in Appendix 2.

Results

Sample and randomization checks

Table 2 summarizes overall descriptive statistics, and Table 3 differentiates the main response variables by treatment
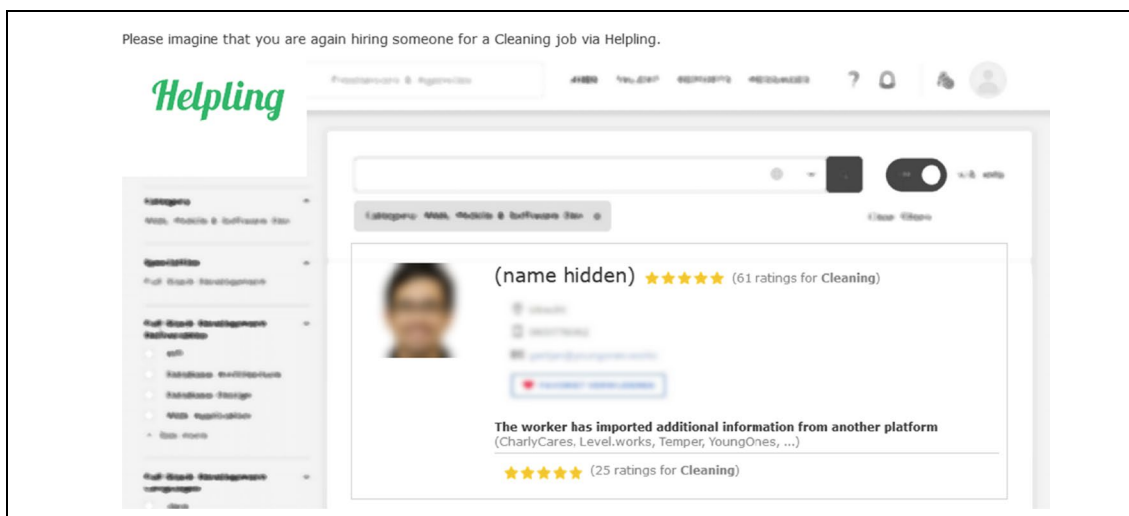
Fig. 3 Stimulus material (example: Helpling; within: same job; imported: same job)

Table 2 Descriptive statistics ($n = 180$)

Variable	Mean	SD	Min	Max
Main response variables				
Trusting belief	4.41	1.68	1	7
Trusting intention	4.22	1.80	1	7
Participant characteristics				
Age	40.89	10.87	18	75
Disposition to Trust				
Gender (male = 0, female = 1)	0.56	-	0	1
has other gender than worker profile	0.03	-	0	1
Uses platform for private purposes	0.56	-	0	1
Uses platform for professional purp.	0.41	-	0	1
Hired more than three times before	0.47	-	0	1
Platforms				
YoungOnes	0.22	-	0	1
CharlyCares	0.20	-	0	1
Helpling	0.41	-	0	1
Temper	0.14	-	0	1
Level.works	0.04	-	0	1
Job Category				
Babysitting	0.20	-	0	1
Hospitality	0.12	-	0	1
Cleaning	0.42	-	0	1
Delivery/logistics	0.16	-	0	1
Promotion	0.02	-	0	1
Retail	0.08	-	0	1

condition. Overall, we see a strong correlation between trusting intention and trusting belief ($r(178) = .87$, $p < 0.001$).

An important feature of any experimental design is that it properly randomizes subjects into treatment groups (Nguyen & Kim, 2019). To check whether this was the case, we ran a set of OLS and logit regressions using the client-specific features: age, disposition to trust, gender, perceived realism, familiarity (with crowd work), and clients' individual importance of workers' reputation, job experience, and platform experience as dependent variables. The independent variables were binary dummies indicating the treatment conditions with regard to the within-platform as well as the imported rating (see treatment matrix above and Table 3). Table 4 shows that, as expected, all variables are statistically indistinguishable across treatment conditions. We hence conclude that randomization was successful.

Treatment effects

Figure 4 shows the average trusting intention and trusting belief by condition, taking into account all the control variables. Moreover, Fig. 5 shows these trust variables as interactions for the contextual signals (within-platform × imported

ratings; top panel) as well as for the contentual signals (same-job × other-job ratings; lower panel). As can be seen there, both signal types have positive effects where within-platform ratings and same-job ratings seem to have somewhat stronger effects than imported and other-job ratings, respectively.

To corroborate this first visual assessment statistically and to test the hypotheses, we conducted a set of OLS regression analyses. We ran the analyses separately for trusting beliefs (Table 5) and for trusting intention (Table 6) as the dependent variable. We z-standardized all continuous independent variables. For all analyses, we controlled for the platform and for job type.[1] We also controlled for participants' age and gender, and we included two dummy variables that indicated whether the participant had used the platform for private and/or professional use. Also, we included a variable that indicated whether participants had hired workers more than 3 times (i.e., median split) in the past. Moreover, we controlled for the worker's gender (as subtly indicated by the blurred profile picture). For readability, we excluded the results for the control variables in Tables 4 and 5. The full tables can be found in Appendix 3.

In support of H1a, within-platform ratings increase trust: clients are more likely to hire workers with a within-platform rating than without, and they perceive them as more trustworthy. The coefficients of "imported rating" (Models 2 and 6) indicate the effect of an imported rating when no onsite rating is present. Here, an imported rating has no significant effect; H1b is thus not supported.

To test H1c, we use a Wald test to compare the coefficients of within-platform and imported ratings. The overall coefficients of within-platform reputation are significantly larger in both Models 1 and 5, that is, for trusting intention ($F = 15.02$, $p < 0.001$) and for trusting belief ($F = 11.13$, $p = 0.001$). H1c is thus supported.

We included the interactions between rating variables in Models 2, 4, 6, and 8. The coefficients of "within platform rating" in Models 2 and 6 hence indicate the effect of within platform ratings when no imported rating is present. The interaction coefficients between within-platform and imported ratings are significant and negative in the model with trusting intention as the dependent variable (Model 2), but insignificant in the model with trusting belief as the dependent variable (Model 6). This means that we find partial support for H1d: the effect of imported ratings on trusting intention is larger when no within-platform ratings are present, while we see no significant interaction for trusting belief.

To test Hypotheses 2a and 2b about the main effects of same-job and other-job ratings, we used an analogous

---

[1] Because the only available job on CharlyCares was babysitting, and because babysitters could only be hired via CharlyCares, we did not include a dummy for babysitting.

Table 3 Summary statistics per condition (mean and standard deviation in parentheses)

			Imported rating			
			None	Other-job	Same-job	Total
Within-platform rating	*N*	**None**	32	26	19	**77**
	Trusting intention		2.97 (1.73)	3.38 (1.90)	4.32 (1.34)	**3.44 (1.77)**
	Trusting belief		3.41 (1.70)	3.73 (1.78)	4.21 (1.32)	**3.71 (1.65)**
	N	**Other-job**	6	7	7	**20**
	Trusting intention		5.33 (1.03)	4.86 (1.57)	5.71 (0.76)	**5.30 (1.17)**
	Trusting belief		5.33 (0.82)	5.29 (1.11)	5.71 (0.76)	**5.45 (0.89)**
	N	**Same-job**	29	32	22	**83**
	Trusting intention		4.79 (1.76)	4.44 (1.56)	4.86 (1.70)	**4.67 (1.66)**
	Trusting belief		4.72 (1.69)	4.84 (1.59)	4.82 (1.56)	**4.80 (1.60)**
	N	**Total**	67	65	48	**180**
	Trusting intention		**3.97 (1.94)**	**4.06 (1.78)**	**4.77 (1.51)**	**4.22 (1.79)**
	Trust belief		**4.15 (1.78)**	**4.45 (1.71)**	**4.71 (1.44)**	**4.41 (1.68)**

procedure. Models 3, 4, 7, and 8 show the results of these regressions. The results support H2a: same job ratings increased trusting intention and trusting belief. Other job ratings only had an effect on trusting intention, but not on trusting belief, so we find partial support for H2b. We also find partial support for H2c: A Wald test shows that the effect of same-job ratings is larger than the effect of other-job ratings ($F = 9.64$, $p = 0.006$) for trusting intention. For trusting belief, the difference is insignificant ($F = 2.97$, $p = 0.087$). The interactions between same job ratings and other job ratings are insignificant in both models (not supporting H2d).

Robustness checks
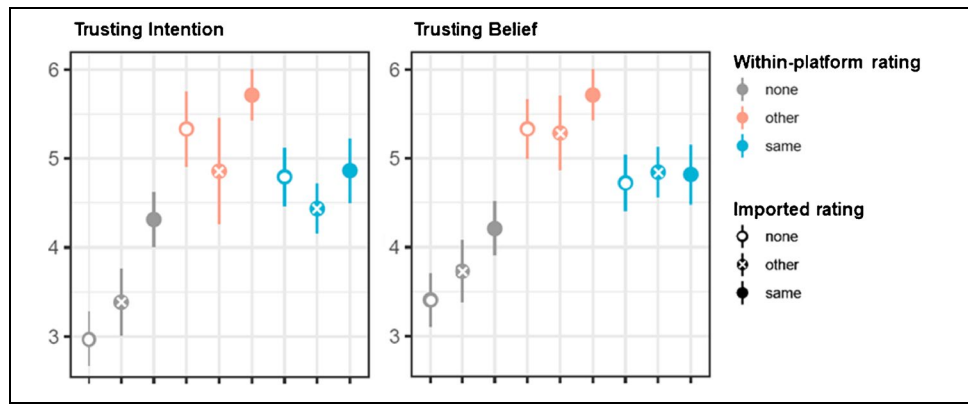
Non-response bias

As we observed a low overall response rate of approximately 2%, we control for non-response bias. To do so, we employ a commonly used technique, comparing the first half of our sample (early responders) with the second half of our sample (late(r) responders). Of course, this is only a proxy that tries to capture potential trends, assuming that response time is associated to responsiveness and

Table 4 Regression models for randomization checks (standard errors in parentheses)

	Participants' …					Importance of …		
	Age	Disposition to trust	Gender ($f = 1$, $m = 0$)	Perceived realism	Familiarity (crowd work)	Reputation	Job experience	Platform experience
Treatment conditions	*OLS*	*OLS*	*logistic*	*OLS*	*OLS*	*OLS*	*OLS*	*OLS*
Has within-platform rating	2.078	.011	−.024	−.163	.126	.389	−.280	.359
	(2.692)	(.240)	(.501)	(.417)	(.441)	(.224)	(.281)	(.334)
Has imported rating	1.160	−.136	−.389	−.230	−.576	.042	.121	.270
	(2.465)	(.219)	(.459)	(.382)	(.404)	(.205)	(.257)	(.306)
Has same-job rating	.183	−.020	.151	.159	−.087	−.413	−.456	−.451
	(3.154)	(.281)	(.591)	(.489)	(.517)	(.262)	(.329)	(.391)
Has other-job rating	−3.252	.280	.789	−.263	.219	−.042	−.036	−.078
	(2.530)	(.225)	(.477)	(.392)	(.414)	(.210)	(.264)	(.314)
Constant	40.026***	5.210***	.126	5.134***	5.045***	6.270***	6.065***	5.171***
	(1.652)	(.147)	(.308)	(.256)	(.271)	(.138)	(.172)	(.205)
Observations	180	180	180	180	180	180	180	180
R^2	.025	.015		.018	.021	.024	.016	.013
Adjusted R^2	.002	−.007		−.005	−.001	.002	−.007	−.010
Log likelihood			−121.474					
Akaike Inf. Crit.			252.948					

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

**Fig. 4** Trusting intention and trusting belief by treatment. Error bars indicate 95% confidence intervals

Job categories

Participants who saw "other job ratings" always saw a worker that earned ratings in a different job than the last job they hired someone for. However, it may be that the specific job was different, but the jobs fell in the same broader category. This was the case for 10 within-platform ratings and 6 imported ratings. We ran a robustness check in which we constructed the other-job variable such that it indicated whether the worker had a rating for a job from a different category. This changes results about other-job rating in the model with trusting intention: the effect of other-job ratings

ultimately does — at least to some degree — reflect differences between responders and non-responders (Koch & Blohm, 2015). Using this approach, we did not find any differences across any of the tested variables (i.e., trust, perceived realism, age, gender, disposition to trust, experience with the respective platform). Moreover, we also compared the first vs. the last *third* of the data: again, no statistical differences in any of these variables. The same holds true when comparing the first vs. the last *quarter* of the data. Overall, we are hence confident that non-response bias does not play an all too severe role for our study – despite the overall low response rate.

Fig. 5 Trusting intention and trusting belief by contextual signal type (onsite/imported) and contentual signal type (same-job/other-job). Error bars indicate 95% confidence intervals
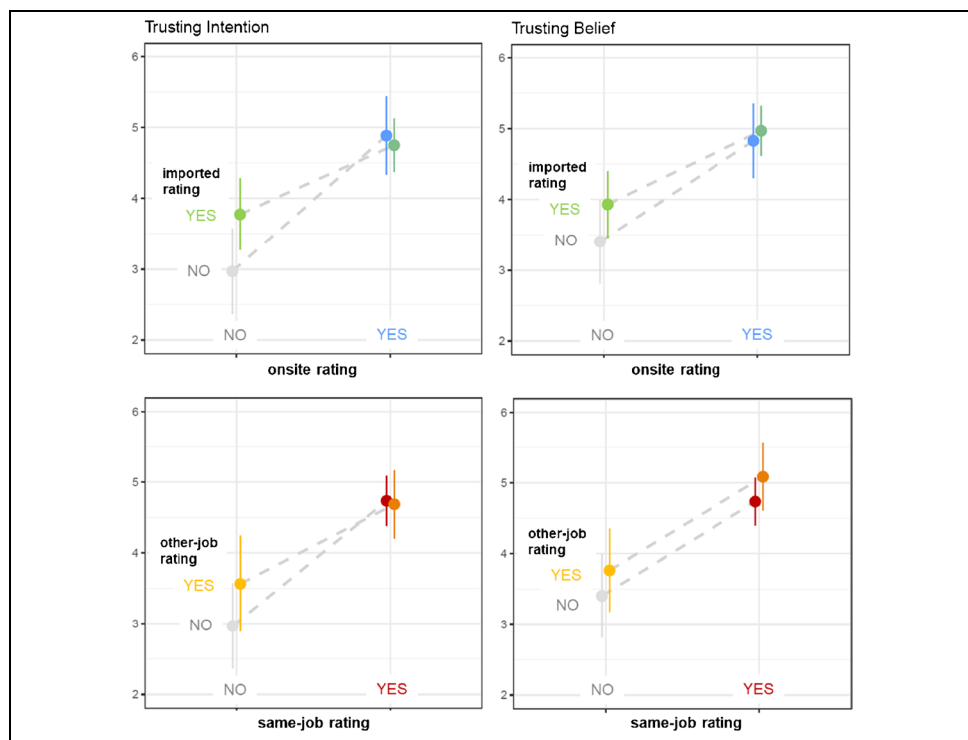
Table 5 Regression results with trusting intention as dependent variable. Control variables omitted here; all coefficients reported in Appendix 3

	Model 1	Model 2	Model 3	Model 4
Intercept	2.41* (1.17)	2.17 (1.17)	1.78 (1.18)	1.58 (1.18)
Main effects				
Within platform rating	1.46*** (0.25)	2.10*** (0.40)	-	-
Imported rating	0.12 (0.25)	0.67 (0.37)	-	-
Same job rating	-	-	1.38*** (0.26)	1.73*** (0.36)
Other job rating	-	-	0.38 (0.25)	0.79* (0.39)
Interactions				
Within × imported	-	−1.03* (0.51)	-	-
Same job × other job	-	-	-	−0.72 (0.52)
R^2	0.298	0.315	0.273	0.282

$N = 180$. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

turns insignificant ($b = 0.47$, $t = 1.14$, $p = 0.26$). There are no substantial changes in the model with trusting belief.

Exploratory analyses

Model variations

We found that only within-platform ratings had an effect on trusting belief and trusting intention and that same-job ratings had a larger effect than other-job ratings. We now explore whether the effect of imported ratings is conditional on the similarity of the job in which it is earned. We do so by re-running Model 2 (Table 5) and Model 6 (Table 6) with the subset of respondents who did not see imported ratings earned in a different job. The effect of imported ratings now reflects the effect of same-job imported ratings. We find that the effect of imported ratings on trusting intention becomes significant ($b = 1.15$, $t = 2.34$, $p = 0.022$), but not significant in the model for trusting belief ($b = 0.55$, $t = 1.17$, $p = 0.246$). The interaction between within-platform

and imported reputation stays significant in the model with trusting intention ($b = −1.41$, $t = −2.05$, $p = 0.044$), but not in the model with trusting belief ($b = −0.65$, $t = −0.99$, $p = 0.233$). When within-platform ratings are present, the effect of imported same-job ratings becomes insignificant ($b = −0.354$, $t = −0.760$, $p = 0.451$). Finally, the difference between the coefficient of within-platform reputation and imported reputation turns insignificant in the model with trusting intention ($F = 3.74$, $p = 0.057$) and the model with trusting belief ($F = 3.75$, $p = 0.056$).

In contrast, when we only include respondents who saw other-job ratings, the effect of imported ratings is insignificant in the trusting intention model ($b = 0.34$, $t = 0.77$, $p = 0.447$) and in the model with trusting belief ($b = 0.21$, $t = 0.42$, $p = 0.615$). The interaction between imported and within-platform reputation was also insignificant in the model with trusting intention ($b = −1.74$, $t = −1.41$, $p = 0.164$) and trusting belief ($b = −0..33$, $t = −0.286$, $p = 0.776$). When only looking at other-job ratings, the coefficient of the difference between within-platform and imported ratings is insignificant in the model with trusting

Table 6 Regression results with the trusting belief in the worker as dependent variable. Control variables are omitted here; all coefficients reported in Appendix 3

	Model 5	Model 6	Model 7	Model 8
Intercept	2.37* (1.09)	2.29* (1.10)	1.84 (1.11)	1.77 (1.12)
Main effects				
Within platform ratings	1.35*** (0.23)	1.57*** (0.38)	-	-
Imported ratings	0.23 (0.24)	0.42 (0.35)	-	-
Same job ratings	-	-	1.13*** (0.25)	1.26*** (0.34)
Other job ratings	-	-	0.56* (0.24)	0.71 (0.37)
Interactions				
Within × imported	-	−0.35 (0.48)	-	-
Same job × other job	-	-	-	−0.27 (0.49)
R^2	0.303	0.305	0.259	0.260

$N = 180$. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

intention ($F = 2.99$, $p = 0.090$) and trusting belief ($F = 1.55$, $p = 0.218$). Together, these findings suggest that imported ratings are associated with trusting intention, but with trusting beliefs, and only when the ratings are earned in the same type of job and when no within-platform ratings are present.

Importance of "within-platform" vs. "same-job" feature

We now explore which feature matters most, that is, whether the same-job property or the within-platform property is more important. This provides an insight for what can be the next best alternative when no within-platform, same-job ratings are available (yet). In other words, is it better to import same-job ratings (from another platform) or to rely on other-job ratings (from the same platform)? A direct approach to this is to compare the respective two treatment cells (see Fig. 2) where only an other-job (within-platform) rating *or* a same-job (imported) rating is available. The relevant cells have $n = 6$ (within: other, imported: none) and $n =1\,9$ (within: none, imported: same) observations, respectively. An independent samples t-test (Welch) shows that trusting beliefs are significantly higher in the within/other (mean = 5.333) than in the imported/same (4.211) category ($p = 0.026$). Similarly, trusting intentions are higher in the within/other (mean = 5.333) than in the imported/same (4.316) category, while this difference is only "marginally" significant at the 10% level ($p = 0.077$). Overall, there hence seems to be a preference towards within-platform ratings (even though earned for different jobs) over imported ratings for the same job type.

Discussion

Summary and contributions

With this study, we aimed to disentangle the trust-building effects of ratings based on their source (within a platform or imported from another platform), and the task they have been earned for. We hypothesized that (positive) ratings imported from a different platform foster trust, but less so than within-platform ratings and less so when within-platform ratings are also present. Likewise, we hypothesized that (positive) ratings about a different job than the focal job foster trust, but less so than ratings about the same job and less so when same-job ratings are present. We tested these hypotheses using an online experiment among actual clients of five gig platforms.

We find that imported ratings may have an effect on trusting intention, but only when the imported ratings are earned with the same job, and only when no within-platform ratings are present. Imported ratings never increased trust *beliefs* about workers. While the effect of imported ratings

was limited, within-platform ratings have a strong positive effect on trusting intention and trust beliefs, in line with the existing literature of reputation effects. Within-platform ratings are thus more strongly associated with trust than ratings imported from a different platform.

Focusing on same-job versus different-job ratings, we find that same-job ratings foster trust in terms of both intentions and beliefs, while other-job ratings only do so for trusting intentions, although this effect is smaller than the effect of same-job ratings. We conclude that, as hypothesized, same-job ratings are more effective than other-job ratings to convey a worker's trustworthiness. We do however not find that the effect of other-job ratings depends on the presence of same-job ratings.

Our findings suggest that while cross-platform ratings *can* foster trust, this effect is at best confined to rather limited conditions, namely, when referring so similar jobs and in the absence of within-platform ratings. Even under those conditions, the effect of imported reputation is qualitatively smaller than the effect of within-platform reputation. This may be explained by differences in the rating system between platforms, which make it difficult for clients to compare within-platform and imported ratings and to assess the usefulness of imported ratings. Future studies may further investigate the relevance of these reasons for the difference in the effect of the different types of ratings.

Our findings on the role of job similarity furthermore suggests that while some information conveyed by ratings can well be transferred to different contexts, there are also specific skills required for jobs that will not transfer well. This may be a more prominent problem on gig platforms than in the sharing economy, since the skills required for different jobs may be more varied than the skills for borrowing or lending goods. Moreover, on platforms for goods-sharing, the owner of the good generally also shows pictures of the good to be shared, while "work" shared on gig-platforms is less-tangible, and therefore riskier. This difference between the gig economy and the sharing economy may also explain why the effect of imported, other-job ratings that we found in the current study is smaller than the effect found in earlier research (Otto et al., 2018; Teubner et al., 2020). Specifically, these studies suggest relative trust increases of 31% (within-platform, same task), 12% (imported, other task), and 26% (imported, other task) — compared to the control condition. In comparison, our findings show that within-platform-, same-job ratings have a larger effect on trusting intention (71%) and trusting belief (45%). Imported, other-job ratings increased trusting intention by 16% and trusting belief by 10%. The effect of imported ratings on trusting intention was thus smaller in our experiment than in previous literature. Also, when comparing the effect of imported, other-job ratings relative to the effect of within-platform, same-job ratings, the effects are smaller than in Teubner

et al. (2020). The difference between the control condition and the imported ratings treatment was 38% of the difference between the control condition and within-platform ratings in their study. In our study, this percentage was 22% for trusting intention, and 23% for trusting belief.

Other explanations for this difference may be that our participants were actual platform clients rather than general Internet users. These real users may have had more experience with assessing workers on the basis of within-platform ratings, and may, as a consequence, be more convinced about the importance of these ratings. Importing ratings however is not a common practice, and imported ratings may therefore be regarded with more caution.

We contribute to the broader theoretical debate on the robustness of reputation signals against various sources of imperfection of information (Bolton et al., 2004, 2005) by identifying contextual and contentual fit as relevant dimensions of such imperfection. Furthermore, we shed light on the empirical conditions that limit this robustness in the context of the gig economy. In contrast to some of theoretical literature (e.g., Nowak & Sigmund, 1998), but in line with earlier empirical findings (Bolton et al., 2004, 2005; Norbutas et al., 2020a), we find that the effectiveness of reputation from a more "distant" source, in our case a different platform, is limited to specific conditions, namely, when there is no within-platform reputation available *and* the task under consideration is similar.

Limitations and future research

We highlight a few limitations of our study and opportunities for further research. First, in order to keep our experimental design parsimonious, we left out many possibly relevant aspects of rating systems. For instance, we only evaluated the source of the ratings, and the job in which the rating was earned; we did not investigate other types of imported reputation, such as automatically generated information (e.g., show-up rate) and other information provided by clients (e.g. written reviews). Likewise, we did not study how the presentation of ratings affects the outcomes. For example, future research could study how average ratings versus distributions of ratings affect trust. Similarly, we only varied *whether* profiles contain certain types of ratings, but not the frequency or the value of such ratings. A promising avenue for future research is to assess under what conditions, in terms of value and frequency,

imported rating capable of outweighing within-platform ratings. Future research could also dive deeper in the importance of specific skills for different jobs. Generating a clear mapping of skills for different jobs would help to compare ratings across jobs and platforms.

Second, while we believe our study improves on previous work by studying a sample of actual gig economy decision makers rather than a convenience sample, our selection of platforms was necessarily restricted to a relatively small set of platforms, with clients from a single country. Future research could aim to generalize our findings to other platforms and a broader set of countries.

Third, although a vignette experiment such as the current study allows for a causal interpretation between the information presented and the stated intentions of the respondents, the setup is also somewhat artificial due to the hypothetical nature of the decisions. A challenge for future research is to develop research designs on reputation portability in more natural settings that nevertheless allow for causal inference. The increasing attention to reputation portability may create new opportunities for such research, for example, in the form of carefully designed field experiments in close collaboration with gig platforms.

Conclusion

Based on the current study, we conclude that imported ratings may help to decrease the lock-in for workers in the gig-economy, but only if the ratings are earned in a similar job as the job a worker is currently applying for. We specifically focused on platforms where clients manually select their worker, rather than platforms where an algorithm assigns workers to clients. It is important to note that while portable reputation may help solve the problem for workers who already have ratings on a different platform for a similar job, they may make it even more difficult for workers who have not used a different platform before, or only for a different type of job. These completely new workers are now competing not only with workers who have previous experience with the current platform but also with workers who have work experience on a different platform. Thus, while in the current study we looked only at individual client behavior, (unintended) market-level effects of these behaviors are an interesting avenue for further research.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quantitative Journal of Economy, 84*, 488–500. https://doi.org/10.2307/1879431

Benard, S. (2013). Reputation systems, aggression, and deterrence in social interaction. *Social Science Research, 42*(1), 230–245. https://doi.org/10.1016/j.ssresearch.2012.09.004

Boero, R., Bravo, G., Castellani, M., & Squazzoni, F. (2009). Reputational cues in repeated trust games. *The Journal of Socio-Economics, 38*(6), 871–877. https://doi.org/10.1016/j.socec.2009.05.004

Bolton, G. E., Breuer, K., Greiner, B., & Ockenfels, A. (2023). Fixing feedback revision rules in online markets. *Journal of Economics & Management Strategy*, jems.12512. https://doi.org/10.1111/jems.12512

Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? *An experimental investigation. Management Science, 50*(11), 1587–1602. https://doi.org/10.1287/mnsc.1030.0199

Bolton, G. E., Katok, E., & Ockenfels, A. (2005). Cooperation among strangers with limited information about reputation. *Journal of Public Economics, 89*(8), 1457–1468. https://doi.org/10.1016/j.jpubeco.2004.03.008

Buskens, V., & Raub, W. (2002). Embedded trust: Control and learning. *Group Cohesion, Trust and Solidarity, 19*, 167–202. https://doi.org/10.1016/S0882-6145(02)19007-2

Buskens, V., & Weesie, J. (2000). An experiment on the effects of embeddedness in trust situations: Buying a used car. *Rationality and Society, 12*(2), 227–253. https://doi.org/10.1177/104346300012002004

Charness, G., Du, N., & Yang, C. L. (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior, 72*(2), 361–375. https://doi.org/10.1016/j.geb.2010.09.002

Cheshire, C. (2007). Selective incentives and generalized information exchange. *Social Psychology Quarterly, 70*(1), 82–100. https://doi.org/10.1177/019027250707000109

Choudary, S. P. (2018). *The architecture of digital labour platforms: Policy recommendations on platform design for worker well-being (No. 3; ILO Future of Work Research Paper Series)*. International Labour Organization.

Ciotti, F., Hornuf, L., & Stenzhorn, E. (2021). *Lock-in effects in online labor markets* (No. 9379; CESifo Working Paper). https://ssrn.com/abstract=3953015. Accessed 28 Apr 2023.

Cook, K. S., Snijders, C., Buskens, V., & Cheshire, C. (Eds.). (2009). *Etrust: Forming relationships in the online world*. Russell Sage Foundation.

Cui, R., Li, J., & Zhang, D. J. (2020). Reducing discrimination with reviews in the sharing economy. *Management Science, 66*(3), 1071–1094. https://doi.org/10.1287/mnsc.2018.3273

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science, 49*(10), 1407–1424. https://doi.org/10.1287/mnsc.49.10.1407.17308

Dellarocas, C. (2010). Online reputation systems: How to design one that does what you need. *MIT Sloan Management Review, 51*(3), 33–38.

Demary, V. (2015). *Competition in the sharing economy*. (No. 19; IW Policy Paper).

DGB. (2021). The German trade union confederation's position on the platform economy. Position paper. https://www.dgb.de/downloadcenter/++co++6a41577e-a1ea-11eb-bae1-001a4a160123. Accessed 28 Apr 2023.

Diekmann, A., Jann, B., Przepiorka, W., & Wehrli, S. (2014). Reputation formation and the evolution of cooperation in anonymous online markets. *American Sociological Review, 79*(1), 65–85. https://doi.org/10.1177/0003122413512316

Duffy, J., Xie, H., & Lee, Y. J. (2013). Social norms, information, and trust among strangers: Theory and evidence. *Economic Theory, 52*(2), 669–708. https://doi.org/10.1007/s00199-011-0659-x

Edelman, B., & Luca, M. (2014). *Digital Discrimination. The Case of Airbnb.com*. https://doi.org/10.2139/ssrn.2377353

European Commission. (2017). Exploratory study of consumer issues in peer-to-peer platform markets. Technical report. https://commission.europa.eu/publications/exploratory-study-consumer-issues-peer-peer-platform-markets_en. Accessed 28 Apr 2023.

Farell, J., & Klemperer, P. (2007). Coordination and lock-in: Competition with switching costs and network effects. In *M. Armstrong & R. Porter (Eds.), Handbook of Industrial Organization*, Elsevier, pp. 1967–2072.

Fehrler, S., & Przepiorka, W. (2013). Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. *Evolution and Human Behavior, 34*(2), 139–145. https://doi.org/10.1016/j.evolhumbehav.2012.11.005

Frey, V., & Van De Rijt, A. (2016). Arbitrary inequality in reputation systems. *Scientific Reports, 6*(1), 1–5. https://doi.org/10.1038/srep38304

Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*, 481–510.

Hesse, M., & Teubner, T. (2020a). Reputation portability–quo vadis? *Electronic Markets, 30*(2), 331–349. https://doi.org/10.1007/s12525-019-00367-6

Hesse, M., & Teubner, T. (2020b). Takeaway trust: A market data perspective on reputation portability in electronic commerce. *HICSS 2020 Proceedings*, 5119–5128

Hesse, M., Lutz, O., Adam, M. T. P., & Teubner, T. (2020). Gazing at the stars: How signal discrepancy affects purchase intentions and cognition. ICIS 2020 Proceedings, pp. 1–9. https://aisel.aisnet.org/icis2020/sharing_economy/sharing_economy/8/

Hesse, M., Teubner, T., & Adam, M. T. P. (2022). In stars we trust—A note on reputation portability between digital platforms. *Business & Information Systems Engineering, 64*(3), 349–358.

Jiao, R., Przepiorka, W., & Buskens, V. (2021). Reputation effects in peer-to-peer online markets: A meta-analysis. *Social*

Science Research, 95, 102522. https://doi.org/10.1016/j.ssresearch.2020.102522

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies, 59*(1), 63–80. https://doi.org/10.2307/2297925

Kas, J., Corten, R., & van de Rijt, A. (2022). The role of reputation systems in digital discrimination. *Socio-Economic Review, 20*(4), 1905–1932. https://doi.org/10.1093/ser/mwab012

Katz, M. L., & Shapiro, C. (1994). Systems competition and network effects. *Journal of Economic Perspectives, 8*(2), 93–115. https://doi.org/10.1257/jep.8.2.93

Koch, A., & Blohm, M. (2015). *Nonresponse bias*. GESIS - Leibniz-Institut für Sozialwissenschafte.

Kokkodis, M. (2021). Dynamic, multidimensional, and skillset-specific reputation systems for online work. *Information Systems Research, 32*(3), 675–1097. https://doi.org/10.1287/isre.2020.0972

Kokkodis, M., & Ipeirotis, P. G. (2016). Reputation transferability in online labor markets. *Management Science, 62*(6), 1687–1706. https://doi.org/10.1287/mnsc.2015.2217

Kollock, P. (1999). The production of trust in online markets. *Advances in Group Processes, 16*(1), 99–123.

Koutsimpogiorgos, N., van Slageren, J., Herrmann, A. M., & Frenken, K. (2020). Conceptualizing the gig economy and its regulatory problems. *Policy & Internet, 12*(4), 525–545. https://doi.org/10.1002/poi3.237

Kroher, M., & Wolbring, T. (2015). Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. *Social Science Research, 53*, 311–324. https://doi.org/10.1016/J.SSRESEARCH.2015.06.003

Lambrecht, C., & Heil, H. (2020). *So schaffen wir die soziale digitale Marktwirtschaft. WirtschaftsWoche*.

Lehdonvirta, V. (2022). *Cloud Empires: How Digital Platforms Are Overtaking the State and How We Can Regain Control*. MIT Press.

Liu, M., Brynjolfsson, E., & Dowlatabadi, J. (2021). Do digital platform reduce moral hazard? The case of Uber and taxis. *Management Science, 67*(8), 4665–4685. https://doi.org/10.1287/mnsc.2020.3721

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 55–63. https://doi.org/10.1126/science.159.3810.56

Nannestad, P. (2008). What have we learned about generalized trust, if anything? *Annual Reviw of Political Science, 11*, 413–436. https://doi.org/10.1146/annurev.polisci.11.060606.135412

Nguyen, Q., & Kim, T. H. (2019). Promoting adoption of management practices from the outside: Insights from a randomized field experiment. *Journal of Operations Management, 65*(1), 48–61. https://doi.org/10.1016/j.jom.2018.11.001

Norbutas, L., Ruiter, S., & Corten, R. (2020a). Believe it when you see it: Dyadic embeddedness and reputation effects on trust in cryptomarkets for illegal drugs. *Social Networks, 63*, 150–161. https://doi.org/10.1016/j.socnet.2020.07.003

Norbutas, L., Ruiter, S., & Corten, R. (2020b). Reputation transferability across contexts: Maintaining cooperation among anonymous cryptomarket actors when moving between markets. *International Journal of Drug Policy, 76*, 102635. https://doi.org/10.1016/j.drugpo.2019.102635

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*(6685), 573–577. https://doi.org/10.1038/31225

Otto, L., Angerer, P., & Zimmerman, S. (2018). Incorporating external trust signals on service sharing platforms. *ECIS Proceedings*.

Przepiorka, W. (2013). Buyers pay for and sellers invest in a good reputation: More evidence from eBay. *The Journal of Socio-Economics, 42*, 31–42. https://doi.org/10.1016/j.socec.2012.11.004

Raad, S.-E. (2020). *Hoe werkt de platformeconomie? (No. 9)*. Sociaal-Economische Raad.

Rani, U., Kumar Dhir, R., Furrer, M., Göbel, N., Moraiti, A., & Cooney, S. (2021). *World employment and social outlook: The role of digital labour platforms in transforming the world of work*. International Labour Organization.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM, 43*(12), 45–48. https://doi.org/10.1145/355112.355122

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. In M. R. Baye (Ed.), *The economics of the Internet and E-commerce. Vol. 11. Adv. Appl. Microeconomics. (pp. 127–157)*. Elsevier Science.

Resnick, P., & Zeckhauser, R. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics, 9*(2), 79–101. https://doi.org/10.1007/s10683-006-4309-2

Rooks, G., Raub, W., & Tazelaar, F. (2006). Ex post problems in buyer-supplier transactions: Effects of transaction characteristics, social embeddedness, and contractual governance. *Journal of Management and Governance, 10*(3), 239–276. https://doi.org/10.1007/s10997-006-9000-7

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

Schoenmueller, V., Netzer, O., & Stahl, F. (2018). *The extreme distribution of online reviews: Prevalence, drivers and implications*. (No. 18–10; Columbia Business School Research Paper).

Schor, J. B., Attwood-Charles, W., Consoy, M., Ladegaard, I., & Wengronowitz, R. (2020). Dependence and precarity in the platform economy. *Theory and Society, 49*(5), 833–861. https://doi.org/10.1007/s11186-020-09408-y

Shaw, J. B. (1990). A cognitive categorization model for the study of intercultural management. *Academy of Management Review, 15*(4), 626–645. https://doi.org/10.5465/amr.1990.4310830

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences, 104*(44), 17435–17440. https://doi.org/10.1073/pnas.0704598104

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics, 8*, 231–340. https://doi.org/10.1146/annurev-economics-080315-015325

ter Huurne, M., Ronteltap, A., Corten, R., & Buskens, V. (2017). Antecedents of trust in the sharing economy: A systematic review. *Journal of Consumer Behaviour, 16*(6), 485–498. https://doi.org/10.1002/cb.1667

ter Huurne, M., Ronteltap, A., Guo, C., Corten, R., Buskens, V., ter Huurne, M., Ronteltap, A., Guo, C., Corten, R., & Buskens, V. (2018). Reputation effects in socially driven sharing economy transactions. *Sustainability, 10*(8), 2674. https://doi.org/10.3390/su10082674

Teubner, T., Adam, M. T. P., & Hawlitschek, F. (2020). Unlocking online reputation: On the effectiveness of cross-platform signaling in the sharing economy. *Business and Information Systems Engineering, 62*(6), 501–513. https://doi.org/10.1007/s12599-019-00620-4

Teubner, T., Hawlitschek, F., & Adam, M. T. P. (2019). Reputation transfer. *Business & Information Systems Engineering, 61*(2), 229–235. https://doi.org/10.1007/s12599-018-00574-z

Teubner, T., Hawlitschek, F., & Dann, D. (2017). Price determinants on Airbnb: How repuation pays off in the sharing economy. *Journal*

of Self-Governance and Management. Economics, 5(4), 53–80. https://doi.org/10.22381/JSME5420173

Tjaden, J. D., Schwemmer, C., & Khadjavi, M. (2018). Ride with me—Ethnic discrimination in Social Markets. *European Sociological Review, 34*(4), 418–432. https://doi.org/10.1093/esr/jcy024

Urzi Brancati, M. C., Pesole, A., & Férnandéz-Macías, E. (2020). *New evidence on platform workers in Europe.* EUR 29958 EN, Publications Office of the European Union, Luxembourg, JRC118570. https://doi.org/10.2760/459278

van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences of the United States of America, 111*(19), 6934–6939. https://doi.org/10.1073/pnas.1316836111

Weigelt, K., & Camerer, C. (1988). Reputation and corporate strategy: A review of recent theory and applications. *Strategic Management Journal, 9*(5), 443–454. https://doi.org/10.1002/smj.4250090505

Zloteanu, M., Harvey, N., Tuckett, D., & Livan, G. (2018). Digital identity: The effect of trust and reputation information on user judgement in the sharing economy. *Plos One, 13*(12), e0209071. https://doi.org/10.1371/journal.pone.0209071