

## ONLINE REPUTATION SYSTEMS AND THE THINNING OF TRUST

*Paolo Parigi & Dan Lainer-Vos\**

### INTRODUCTION

Trust, the skillful suspension of doubt, plays a crucial role in social life and online markets.<sup>1</sup> Two-sided marketplaces, in which two sets of agents exchange goods or services through an online intermediary platform, depend on trust cultivation among strangers.<sup>2</sup> eBay, Uber, Lyft, Airbnb, and other operators of two-sided market platforms rely on buyers and sellers trusting the information each provides regarding the payment, quality, safety, performance of the advertised product or services.<sup>3</sup>

In the absence of product standardization or top-down sanctioning of defectors, reputation is the key mechanism that generates trust between a buyer and a seller.<sup>4</sup> The platforms that operate in online markets build sophisticated reputation systems to facilitate commerce. Online reputation systems consist of two parts: reviews and a set of ratings. Reputation systems differ in the prominence they give to one piece instead of the other and in how they display ratings.

---

\* Paolo Parigi, Associate Director, Computational Social Science at IRiSS and Researcher, Stanford University & Facebook; Dan Lainer-Vos, Professor, Department of Sociology, USC.

<sup>1</sup> RACHEL BOTSMAN, *WHAT'S MINE IS YOURS: THE RISE OF COLLABORATIVE CONSUMPTION* (1st ed., 2010).

<sup>2</sup> Scholars use the term sharing economy or gig economy to describe this phenomenon. We reject the idealized understanding conveyed by the term “sharing economy,” and since our focus is on exchange and trust relations and not on production, we opt to describe this new arena using the term “market” rather than economy.

<sup>3</sup> ARUN SUNDARARAJAN, *THE SHARING ECONOMY: THE END OF EMPLOYMENT AND THE RISE OF CROWD-BASED CAPITALISM* (2016).

<sup>4</sup> Bruno Abrahao et al., *Reputation Offsets Trust Judgments Based on Social Biases Among Airbnb Users*, 114 *PROC. NAT'L ACAD. OF SCIS.* 9848 (2017).

Nevertheless, reputation systems' ubiquity attests to their critical role in fostering trust in two-sided online markets.<sup>5</sup> Reputation facilitates trust between the two sides, and trust underpins these markets' functioning.

The development of two-sided markets and sophisticated reputation systems fundamentally alters the nature of trust. In traditional markets, trust is a cumulative byproduct of repeated dyadic exchanges. Trust is different from blind faith or purely calculated decision. It requires a skillful suspension of doubt on the basis of limited information. As partners engage in repeated face-to-face transactions, mutual trust emerges. In these simple markets, trust was an interactional accomplishment. In such markets, inefficiencies in the diffusion of information constrain the exchange scale.

On the other hand, in two-sided online markets, face-to-face interactions are non-existent or minimal, and trust relations are technologically-mediated.<sup>6</sup> Furthermore, trust is not a cumulative result of repeated dyadic interactions in two-sided markets but an accretive product of the crowdsourced reviews generated by previous sellers and buyers. The technological mediation of two-sided markets enables rapid and efficient diffusion of information and fosters exchange relations at a previously unimaginable scale.<sup>7</sup>

The emergence of two-sided markets also alters the meaning of trust. We refer to this development as the *thinning* of trust. Whereas previously, sellers' trustworthiness took time to cultivate

---

<sup>5</sup> Paul Resnick & Richard Zeckhauser, *Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System*, in THE ECONOMICS OF THE INTERNET AND E-COMMERCE (ADVANCES IN APPLIED MICROECONOMICS) (2002).

<sup>6</sup> MICHAEL MUNGER, TOMORROW 3.0: TRANSACTION COSTS AND THE SHARING ECONOMY (2017).

<sup>7</sup> SUNDARARAJAN, *supra* note 3.

because it depended on person-to-person interactions, now its growth is outsourced to the platform. Trust in traditional markets is mostly personal. It is associated with an individual partner to exchange, and it operates across domains. In contrast, trust in two-sided online platforms is impersonal because the judgment of trustworthiness is relative. We determine that seller X is trustworthy based on the aggregate assessment of others and relative to the judgment of many others on the reliability of other sellers of the same good or service.<sup>8</sup> As a consequence of the impersonal and technologically-driven nature of two-sided markets, the trust that underpins them is *thin* and confined to specific domains such as traveling or lodging.

While online reputation systems have made it easier to trust strangers and facilitate the circulation of trust, they have also removed part of the process we used to learn about one another. This disenchantment is reminiscent of Max Weber's argument on the rationalization of religion at the dawn of modernity.<sup>9</sup> Weber argued that Protestantism rationalized religion by eviscerating magic from religious practice and life. Similarly, we argue that the rise of two-sided markets and the centrality of reputation systems had led to the disenchantment of trust.

Instead of trust being a spontaneous byproduct of interpersonal interaction, thin trust in two-sided online markets demands methodical cultivation,<sup>10</sup> is mostly impersonal, and is

---

<sup>8</sup> 93% of positive reviews for a vendor on Amazon or eBay translates to "trustworthy" only if other vendors on the same page have similar or lower ratings.

<sup>9</sup> MAX WEBER, *THE PROTESTANT ETHIC AND THE SPIRIT OF CAPITALISM WITH OTHER WRITINGS ON THE RISE OF THE WEST* (2008).

<sup>10</sup> That thin trust requires methodical cultivation is evident in the proliferation of reputation management services that help vendors sustain their reputation in a volatile online environment.

domain- specific.

The thinning of trust provides a vivid illustration of the opportunities, risks, and responsibilities that today's social sciences face. In the past, social scientists were, to the most part, observers of society, and their tools had little impact on social organization.<sup>11</sup> The migration of much of our social life to a digital interface, the fact that we shop online, for instance, creates an explosion of data, and more importantly, an exponential growth in our ability to intervene, manipulate, and curate social relations. A small change in the organization of a reputation system, for instance, can be extraordinarily consequential to the parties involved in the exchange. The capacity to measure and intervene is not limited to the domain of trust. Rather it impacts many domains of social life from the most intimate (dating and romantic relationships) to the more collective and discrete (neighbor relations). The process that disenchanting trust is, in other words, repeating itself across all the domains that technology is making measurable.

Currently, platforms leave the design of the techniques that facilitate human interactions in the hands of engineers and designers. Yet, social science today can be practical like never before.

Academic institutions and researchers have only started to grapple with the implications of this new reality. By focusing on the role of the reputation system in creating thin trust, this paper offers a fresh perspective for a new and more applied role of social sciences. We will get back to this point in the conclusions.

---

<sup>11</sup> The distinction we make is relative. We do not claim that social science had no impact or that the tools at our disposal were free of consequences. Surveys, for instance, are extremely consequential (cite). Yet the migration of our significant portion of social life to digital interface, created an explosion of data and an exponential growth in the ability to curate social realities with the strike of a keyboard.

We organize this chapter as follows. First, we will review the relevant literature on the topic of trust. The focus is on interpersonal trust rather than trust in institutions or generalized trust in others. We will then present evidence of how the reputation system is creating trust on these platforms. Most of the evidence comes from Parigi's previous work on platforms like Airbnb, Uber, and CouchSurfing. After presenting this evidence, we will explore the transformation that the concept of trust had undergone in terms of rationalization and examine the opportunities and risks that this new format of trust creates from social researchers and society. In conclusion, we will suggest a novel approach for scientific knowledge in the social sciences that aims at becoming applied.

#### **TRUST IN NETWORKS**

What is a trust? What does it mean to trust another person? Trust hovers between calculated action and blind faith. As Giddens notes, trust is necessary only in conditions of *incomplete* information, where the limits of existing knowledge and calculation, requires actors to suspend their disbelief and commit to a line of action that is inherently risky.<sup>12</sup> (Giddens 1990, 33). The partner to exchange, even one that has proved reliable in prior interaction, can always renege at the last minute. To complete such a transaction, the actors must, at some point, suspend their doubt and commit to the exchange. But the suspension of doubt is not equal to blind faith. In typical circumstances, the trusting actors rely on contextual cues (past experience, the context of the interaction, and the assessment of third parties' behavior) that turn trust into a reasonable if not fully calculated choice.<sup>13</sup>

---

<sup>12</sup> ANTHONY GIDDENS, *THE CONSEQUENCES OF MODERNITY* 33 (1990).

<sup>13</sup> GIL EYAL, *THE CRISIS OF EXPERTISE* (2019).

Given its centrality in social life, social scientists have studied and developed competing conceptions of interpersonal trust.<sup>14</sup> For rational choice theorists, people trust each other because of the benefits that trust generates.<sup>15</sup> Building on this approach while dispensing with the criticism that rationality requires perfect information,<sup>16</sup> some scholars have argued that trust emerges when the interests of the two parties engaged in the interaction are aligned.<sup>17</sup> On the contrary, other scholars have argued that trust is precisely needed when the parties' interests are unknown.<sup>18</sup> Finally, for students of culture, trust between people is the result of norms that shape society and get passed to individuals through institutions like the family and school.<sup>19</sup>

Mark Granovetter places the study of trust on empirical grounds by linking it to concrete social networks: "You may trust that potential leader if there is a link or short chain of personal links to that person that conveys enough information to afford you some confidence that she will act in a trustworthy manner."<sup>20</sup>

Granovetter's work clarifies that trust rests on the flow of information. Yet, he limits this flow to personal links and implicitly equates trust with in-person interactions. Personal links carry information and accountability and, because of that, can create a trust chain.<sup>21</sup> Trust networks in two-sided markets have a different structure than traditional trust networks (see Figure 1). In traditional

---

<sup>14</sup> Karen S. Cook & Bogdan State, *Trust and Economic Organization*, in *EMERGING TRENDS IN THE SOCIAL AND BEHAVIORAL SCIENCES 1* (2015).

<sup>15</sup> JAMES S. COLEMAN, *FOUNDATIONS OF SOCIAL THEORY* (1994).

<sup>16</sup> ROBERT GIBBONS, *A PRIMER IN GAME THEORY* (1992).

<sup>17</sup> RUSSELL HARDIN, *TRUST AND TRUSTWORTHINESS* (2002).

<sup>18</sup> PETER BLAU, *EXCHANGE AND POWER IN SOCIAL LIFE* (1986).

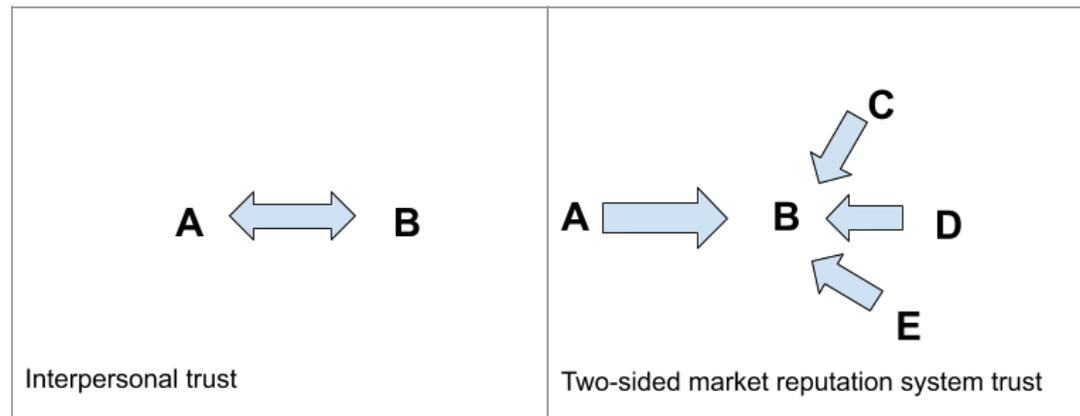
<sup>19</sup> FRANCIS FUKUYAMA, *TRUST: THE SOCIAL VIRTUES AND THE CREATION OF PROSPERITY* (1995).

<sup>20</sup> MARK GRANOVETTER, *SOCIETY AND ECONOMY: FRAMEWORK AND PRINCIPLES* (2017).

<sup>21</sup> Ferdinand Tönnies, *Community and Civil Society* (Margaret Hollis trans., 2001).

settings, trust results from repeated dyadic interaction between individuals (see left panel of Figure 1). Trust builds up incrementally, over time, and solidifies with the completion of each transaction. The repeated interpersonal nature of these transactions means that actors typically attribute trustworthiness to the actors they engage with. After a series of exchanges between A and B, A is likely to assume that B is trustworthy.

*Figure 1: Configurations of trust*



In two-sided markets actors are less likely to engage in repeated transactions. The information that actor A uses to determine the trustworthiness of actor B includes information posted by other actors (C, D, E, etc.) based on *their* previous interaction with B and depends on A's valuation of the quality of the reputation system. Efficient information flow in the reputation system means that trust between strangers emerges almost instantaneously but results from the gradual accumulation of reviews of previous transactions.

Trustworthiness in this situation is impersonal in three senses. First, the information conveyed by previous reviews is not strictly related to B but to the interaction between B and C, B and D, B and E, etc. A must decide whether C's review of B is credible and whether it reflects B or C or something in between. Second, the aggregation of past reviews through star ratings or otherwise means

that when A determines that B is trustworthy or not, it does so by comparing B's ratings to other vendors on the platform. Trustworthiness thus becomes a relative property. In contrast with personal trust, which involves the presumption of reliability and involves the attribution of probity or honor,<sup>22</sup> trust in two-sided market emerges from comparing the ratings of many others. Third, given that trustworthiness of an actor on the platform is tied with the credibility of the reputation system and the platform as a whole, trustworthiness is distributed between the two. Finally, given that two-side markets are domain-specific (Uber provides transportation, AirBnB provides hospitality), an actor's trustworthiness in a given market is not easily transferable to other domains. This is why we call this new type of trust, "thin trust." Thin trust is impersonal and domain-specific.<sup>23</sup> It can effectively connect strangers and facilitate their interaction, but it is much narrower in meaning and scope than interpersonal trust.

The emergence of vast two-sided markets, in which trust relations connect a multiplicity of strangers, requires us to update Granovetter's perspective. We hold on to Granovetter's focus on information flows but note that reputation systems facilitate the flow of massive amounts of personal data between agents and that this information flow sustains expansive trust networks decoupled from personal links. Knowledge of the most arcane things is now just a click away. In personal interactions, reputation systems have

---

<sup>22</sup> Giddens notes that this attribution renders trust psychologically consequential to the individual who trusts. See ANTHONY GIDDENS, *THE CONSEQUENCES OF MODERNITY* (1990).

<sup>23</sup> Unlike trust that is the result from the attribution of probity associated with a specific person, which is not strictly associated with particular domains, the attribution of trust in two-sided market platforms is associated with the particular service one offers. This domain-specific limits is, in part, a direct consequence of the fact that typical two-sided markets are specialized. Uber, for instance, provides transportation services, Airbnb specializes in hospitality, etc. and the reputations aggregated in one platform are not available in another.

distilled detailed and personal information in a digestible way designed for scaling and diffusion. Personal information used to require time to acquire. Now personal information about perfect strangers is available to participants of many online platforms as soon as they join the platform. Updating Granovetter's approach to incorporate technology means extending personal links to encompass online personal ties/information.

Online trust networks differ from Granovetter's face-to-face trust networks not only in scale and speed but also in their intermediation. Unlike spontaneously emerging face-to-face trust, trust in two-sided markets, Trust and Safety teams or divisions within many platforms, cultivate and curate the emerging networks. These teams' objective is to protect their users, but their main byproduct is trust. From this perspective, trust online is a network good that is actively being generated by users of a platform when they contribute with reviews and ratings and tech companies' workers when they make sure the feedback is authentic.<sup>24</sup>

For the most part, the main scope of Trust and Safety teams is to reduce fraud. Platforms attract many scammers who seek to exploit the vulnerabilities of a system that relies on mass participation is loosely supervised. Scams range from a host falsely advertising a property on Airbnb they do not own to elaborate fake accounts on Uber generating demands for reimbursement. Scammers are continually testing the network for vulnerabilities.

---

<sup>24</sup> While it may be the case that reputation systems online emerged by happenstance, their maintenance, control and evolution are essential part of two-sided markets. "When eBay launched, the biggest challenge was that consumers simply did not trust that they would get what they paid for. eBay quickly realized that without consumer trust, the system could not work. In response, eBay created the first Trust and Safety team, which was tasked with ensuring the trustworthiness of the eBay ecosystem." Amy J. Schmitz & Colin Rule, *Lessons Learned on eBay*, in *THE NEW HANDSHAKE: ONLINE DISPUTE RESOLUTION AND THE FUTURE OF CONSUMER PROTECTION* 33 (2017).

Containing fraud and malfeasance behavior is the primary goal of Trust and Safety teams in all two-sided platforms.

The containment and reduction of bad actors increase the reputation system's credibility and contribute indirectly towards increasing trust. The creation of badges and special statuses for users that passed carefully chosen milestones are also essential signals that facilitate trust between parties. An apt example is Airbnb's Super Host status on its platform. Airbnb reserves the status for hosts that meet specific criteria, and it is a signal of the host's trustworthiness. Interventions of this type directly create trust by extending the reputation of users. Thus, trust and safety teams indirectly create trust by containing bad actors and directly creating trust by expanding reputation signals.<sup>25</sup>

#### **HOW ONLINE REPUTATION SYSTEMS GENERATE TRUST**

The nature of online trust networks and their malleability creates opportunities for studies demonstrating what "thin trust" is all about. In a recent study, Parigi and his colleagues experimented to explore the extent to which the reputation system extended trust beyond homophily.<sup>26</sup> Homophily, one of the few constant behaviors of social life,<sup>27</sup> is the tendency to interact and trust others who are similar.<sup>28</sup> The researchers set up an online experiment based on the widely used investment game that simulates actors' behavior in two-sided markets. In two-sided markets, participants decide whom to trust based on the information displayed about the unknown alter.

---

<sup>25</sup> Notice that the expansion of signals is limited to a specific platform. The Super Host status does not apply to a driver for Lyft, for instance.

<sup>26</sup> Abrahao et al., *supra* note 4.

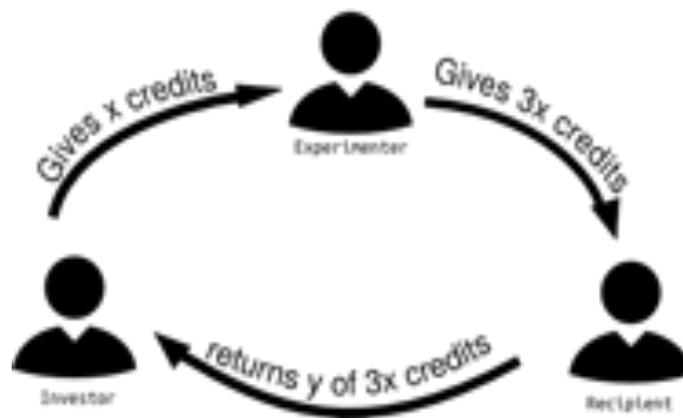
<sup>27</sup> PETER BLAU, *INEQUALITY AND HETEROGENEITY* (1977).

<sup>28</sup> Miller McPherson, *An Ecology of Affiliation*, 48 *AM. SOCIOLOGICAL REV.* 519 (1983).

Similarly, in the investment game, users have to decide whom to place trust based on limited information.

The investment game is a single-shot game where participants decide how many credits to invest in a recipient. Recipients receive three times that amount and may cooperate or defect when determining how many credits to return to the investor. The figure below shows a stylized version of the game:<sup>29</sup>

Figure 2: The standard trust game.



For example, if a participant decided to invest 5 points, the recipient will receive 15 points and choose how many points to return. Parigi and his colleagues led all participants in the experiment to believe that they were randomly assigned the role of investor and instructed them to play with five other Airbnb users cast to recipients' role. In reality, the recipients were synthetic profiles that the researchers concocted. As investors, participants received 100 points and had to decide how to allocate them. The experiment involved almost 9,000 Airbnb users in the United States.<sup>30</sup>

<sup>29</sup> Will Qiu et al., "More Stars or More Reviews?", in PROCEEDINGS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (2018).

<sup>30</sup> Abrahao et al., *supra* note 4.

The synthetic profiles had different demographic characteristics (gender, age, marital status, and U.S. region) and different reputation levels, all varied in a structured way. For the demographic characteristics, profiles were located at various social distances ranging from matching all the participant attributes to differing in all the attributes. A profile at (social) distance 0 had demographic characteristics that fit the participant's profile, while a profile at a distance of 4 was the most dissimilar. To illustrate, imagine a male player from California, not married, and 40 years old. The profile at distance 0 will have all the same characteristics of the player, while the profile at a distance of 4 will be all different, i.e., a female from New York, married in her 60s.

The 5<sup>th</sup> profile was identical to the profile at a distance of 4 but had a different reputation from all others. The experiment had two conditions—one in which the 5<sup>th</sup> profile had a worse reputation than the previous four (world 1, Figure 2 left panel) and one in which she had a better reputation (world 2, Figure 2 right panel).

*Figure 3: Main results of an online experiment*

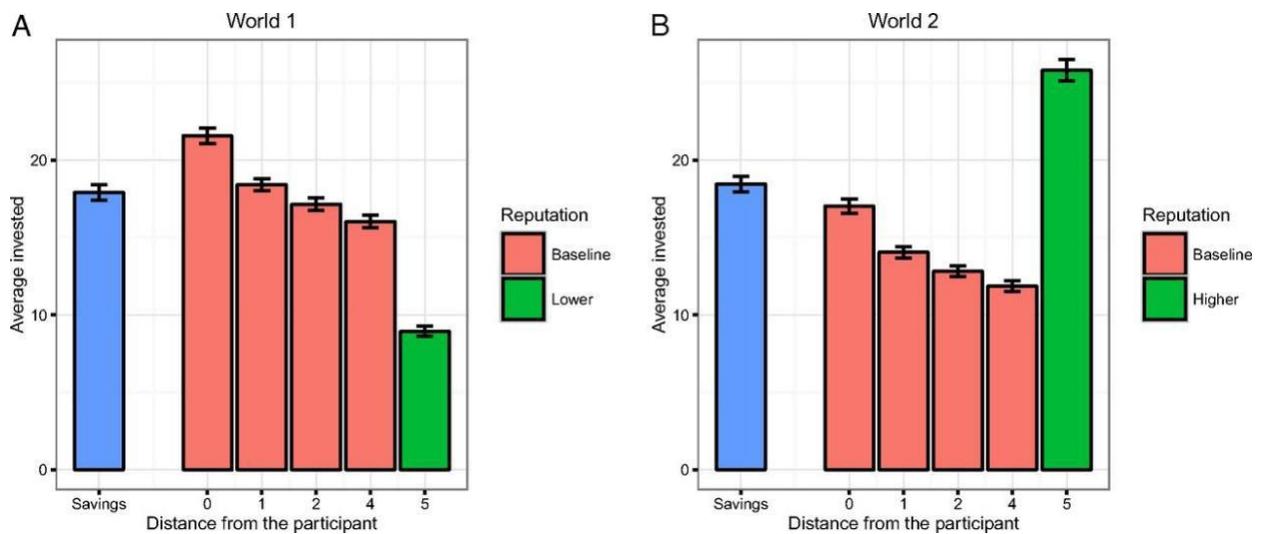


Figure 3 presents the main result of the research. The x-axis plots the social distances on the profiles, 0 to 5. As previously explained, distance 0 means a profile that is identical to the participant; distance 1 means a profile with one characteristic different from the participant, and so on. Distance 5 (green bar) represents a different profile from the participant and has the same characteristics as distance four but a different reputation. Note that there is no distance 3 because of the difficulty in interpreting 3-way interaction effects. The y-axis plots the average amount of points invested. In both panels, the blue bar shows the average amount of points not invested (, i.e., saved).

On the left panel, the effect of homophily is almost self-explanatory. The more socially distant a profile was from the participant, the lower the number of points invested. In other words, the left panel confirms that people trust others who are similar to themselves. Homophily works online as it works offline. Note in this condition, world 1, that the 5<sup>th</sup> profile paid an extra penalty caused by his worse reputation—the decrease in points invested in him compared to the profile at a distance 4 is large and significant.

Focusing on the right panel, or the condition in which the most diverse profile (green bar) has a better reputation than all the other profiles. The plot shows a dramatic increase in trust. A positive reputation significantly extends trust beyond the effects of homophily (still visible in the declining trust in the red bars as social distance increases). After controlling for various factors and considering the complex experiment's dependencies, the researchers concluded that the reputation system significantly extended trust towards different others.<sup>31</sup> The reputation system makes possible the circulation of trust in two-sided markets.

---

<sup>31</sup> *Id.*

Participants/investors interpret the ratings and reviews as signals for trustworthiness, and because of that, engage in the exchanges the platform offers.

### LOOKING AT THE STARS

In most platforms, reputation systems have two components, ratings, and reviews. Ratings are usually expressed on a 5-star scale, while reviews consist of comments that users left about their experiences. Using the same data described above, Qiu et al.<sup>32</sup> separated the impact of the two parts of the reputation system on perceptions of trustworthiness.

In the experiment above, participants were exposed to a star condition—4 or 5-star ratings—and a review condition—a low number of reviews (1-3) or a high number of reviews (11-50). While the difference between 4 stars to 5 stars may appear limited, at first sight, it mirrors a reality in which the overwhelming majority of ratings available on two-sided markets are positive.

Qiu et al. compared the reputation system's impact by focusing on profiles at distance 4 and distance 5. Both profiles have the same demographic characteristics but a different reputation. In particular, they fit the following model:

$$[1] Y_{ij} = \mu + \alpha_j + \beta_1 s_i + \beta_2 r_i + \beta_3 (s_i \times r_i) + \beta_4 w_i + e_{ij}$$

Where  $Y_{ij}$  is the predicted investment amount for profile  $i$  by subject  $j$ .  $\mu$  is the global intercept at a star rating of 4 and Low Review count in world 1, and  $\alpha_j$  are random intercepts to account for individual variations.  $\beta_1$  is the profile level estimate of having 5 stars,  $\beta_2$  is the profile level estimate of having High review counts,

---

<sup>32</sup> Qiu et al., *supra* note 29.

and  $\beta_3$  is their estimated interaction effect.  $\beta_4$  is the estimate of a profile  $i$  being placed in world two, and  $e_{ij}$  is the random error.

Star rating (s) is a factor variable with two levels (4 stars or 5 stars), and review condition (r) also has two levels (Low or High). Because there were multiple measurements of investments per participant (each participant invested p4 and p5), the measured investments are correlated. To account for this, the model nested profile investments within subjects by fitting simple random subject-level intercept  $\alpha_j$ . The model also does not include an explicit term for social distance because we confined our analysis to observed investments between p4 and p5 only, who share the same distance to the subject but different reputations.

Table 1 summarizes the results from Qiu et al. The table shows estimated fixed effect coefficients for five models: (1) an intercept only model, (2) star rating only model, (3) review count only model, (4) additive model of star and review, and lastly a full model with (5) both additive as well as interaction terms.

*Table 1: Multi-level Model Estimates of Star Ratings and Review Counts Components on Investment*

Covariate	Dependent variable:				
	Investment				
	(1) (Intercept Only)	(2) (Star Only)	(3) (Review Only)	(4) (No Interaction)	(5) (Full Model)
Intercept	16.09***(0.1609)	13.16***(0.237)	12.91***(0.235)	9.534***(0.292)	7.08***(0.3718)
Star = 5		5.285*** (0.3187)		5.69***(0.31)	4.497***(0.452)
Review = H			5.799***(0.317)	6.21***(0.311)	5.16***(0.45)
Star = 5 × Review = H					1.487(0.60)
World = two					6.53*** (0.303)
<b>Variance Components</b>					
Subject	0.0	$7.56 \times 10^{-12}$	$3.899 \times 10^{-10}$	14.06	1.58
Residual	226.7	219.8	218.4	197.34	198.94
Observations	8760	8760	8760	8760	8760
AIC	72378.0	72109.4	72052.2	71749.4	71309.8
BIC	72399.3	72137.7	72080.5	71784.8	71359.3
Log Likelihood	-36050.7	-40112.3	-36022.1	-35869.7	-35647.9

The results show a significant increase in average investment received when a profile goes from having 4 stars to 5 stars (~ 4.5 more credits)) as well as going from low review to high review (~ 5.16 more credits). Their interaction ( $\beta_3 = 1.487$ ) is not statistically significant. Having a 5- star rating and lots of reviews does not significantly increase trust in the profile. Either one of the two conditions suffice. Finally, the estimates are stable even when we consider world differences. The researchers summarized their findings: “for a profile, the effect of going from having 4 stars to 5 stars on the number of credits is equivalent to the effect of going from having only 1-3 reviews to having at least 11 reviews on average.”<sup>33</sup>

The arbitrariness of using a 5-star scale for ratings and the peculiarities of many of the comments left on these platforms has made many observers think that the reputation system is an ancillary add-on to many websites. This analysis shows that the reputation system is crucial for creating extended trust. These platforms’ users interpret both the ratings and the reviews as signals for trustworthiness. These signals represent thin trust because the judgment of trustworthiness is both numerical and relative. Users do not engage with the profile characteristics but with the profile in relation to other profiles.

#### **MODELING AND THINNING TRUST**

The capacity to measure trust has been the holy grail of trust scholars for many years because trust is essential for economic growth,<sup>34</sup> the health of institutions,<sup>35</sup> and individual well-being.<sup>36</sup>

---

<sup>33</sup> Qiu et al., *supra* note 29, at 7.

<sup>34</sup> Stephen Knack & Philip Keefer, *Does Social Capital Have an Economic Payoff? A Cross-Country Investigation*, 112 THE Q.J. ECON. 1251 (1997).

<sup>35</sup> John F. Helliwell & Robert D. Putnam, *Economic Growth and Social Capital in Italy*, 21 EASTERN ECON. J. 295 (1995).

<sup>36</sup> ERIC M. USLANER, *THE MORAL FOUNDATIONS OF TRUST* (2002).

Yet, measuring trust has proven elusive due to the concept's subjective and fleeting nature. More importantly, even when researchers successfully measure trust, the techniques for doing so—detailed attitudinal surveys—are imprecise, costly, and therefore quite rare.

The penetration of technology in many aspects of life changed this and made it possible to accumulate data on private interactions that were previously unthinkable. The optimization of the virtual spaces where these interactions occur allows measuring trust using behavioral data rather than relying mostly on costly attitudinal surveys. To the extent that technology has entered many more contexts of contemporary life, from walking your pet, to hosting people, to suggesting potential romantic partners, it has created a world that is amenable to digital experimentation, measurement, and optimization.<sup>37</sup> While trust in your loved one may not be measurable, the trust that circulates on a platform like Airbnb is. A progressively more digital world is also a more quantifiable world. It is also a world where trust can be carefully designed.

Barbosa et al.'s work illustrates these new opportunities.<sup>38</sup> The researchers developed a data triangulation process by which they collected data first using an online experiment very similar to what we described above. The experiment provided them with a measure of the trusting behavior of about 5,000 Airbnb users. Using machine learning, they then create a model to identify low, middle, and high trust levels. The model identified actions taken on the platform, i.e., logged behavior, that correlated with trust levels.

---

<sup>37</sup> Xiao Ma et al., *When Do People Trust Their Social Groups?*, in PROCEEDINGS OF THE 2019 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (2019).

<sup>38</sup> Natã M. Barbosa et al., *Designing for Trust: A Behavioral Framework for Sharing Economy Platforms*, in PROCEEDINGS OF THE WEB CONFERENCE 2133 (2020).

Among the actions that correlate positively with trust among guests were: (a) reading the full description of the host profile and (b) the length of the communication with the host. Factors that were negatively associated were: (a) reading reviews in full and (b) level of engagement with other guests generated content.<sup>39</sup> Among hosts, the positively correlated actions were: (a) total engagement (i.e., time spent in the app), (b) prior requests of engagement, while the negatively correlated behaviors were: (a) the number of communication exchanges and (b) number of rejections.<sup>40</sup>

In a second step, Barbosa et al. validated their trust model by changing the target variable. That is, they used the model to predict answers to a set of trust questions provided by Airbnb hosts and guests. This sample was much larger, about 200,000 respondents, and completely independent of the sample that participated in the experiment. Most relevant, the dependent variables for the predictions were survey questions rather than behavioral variables. The table below summarizes the results of this second step for Airbnb hosts:

*Table 2: Triangulation of attitudes and experimental data.*

<b>Attitudinal Survey Question</b>	<b>Avg. Prediction</b>
<i>How trustworthy are Airbnb guests? (N=4,449)</i>	
High (4-5)	2.14
Low (1-2)	2.07
<i>How safe do you feel when hosting guests in your listing(s) with Airbnb? (N=52,141)</i>	
High (4-5)	2.16
Low (1-2)	2.06
<i>Trust in Airbnb if things go wrong (N=65,317)</i>	
High (4-5)	2.14
Low (1-2)	2.07

<sup>39</sup> *Id.* at 2138.

<sup>40</sup> *Id.*

For all the questions, the high trust group model's average predictions are higher than for the low trust group, and the differences are all statistically significant. In other words, the model did a good job in predicting trust beliefs of users from a different sample. This step connected the behavioral model to the attitudinal model. Guests that exhibited a particular set of behavior correlated with, say, low levels of trust also had low trust beliefs towards hosts. Finally, Barbosa et al. trained a neural network to predict low, medium, and high trust levels for hosts and guests. They applied the neural network model to classify another batch of hosts and guests.

Their model performs very well in predicting users with low levels of trust, while it is less accurate for users with high trust levels.<sup>41</sup>

The procedure we describe above illustrates the technological capacity to measure and intervene in the production of trust and its subsequent thinning. The measurement of trust, which used to be costly and rare, is now accomplished at a relatively low cost and applied to the entire population of users.<sup>42</sup> Once done, the researchers can intervene and study how various changes to the platform change trusting behavior, again, at a scale previously unimaginable. Note that the trust we enact and work on in this setting is thin. Barbosa and his colleagues do not attempt to model and work upon a thick interpersonal relation but merely to alter user behavior on a single platform.

The measurement and thinning of trust rely on the reputation system's technology. In the absence of such technology, trust

---

<sup>41</sup> Barbosa et al., *supra* note 35.

<sup>42</sup> In the past, measurement of trust relied primarily on attitudinal surveys. Such surveys were able to measure general tendency to trust others, for instance. But attitudinal surveys are costly, rare, and lack the specificity required to intervene in practical action.

remains mostly inoperative. One of the early two-sided marketplace platforms, CouchSurfing,<sup>43</sup> is a case in point because it never implemented a reputation system. Instead, it relied heavily on its users posting detailed descriptions of themselves and their interactions. In describing the world of CouchSurfing, Patricia Marx—a writer for the *New Yorker*—wrote:

Upon joining CouchSurfing, you are instructed to compose an online profile, delineating your philosophy and mission, the skills you can teach others, your favorite music, movies, and books, and so much else that you might as well be applying to college.

Members also post photographs of themselves, sometimes hundreds of them.<sup>44</sup> Without the reputation system's technology, trust on the CouchSurfing platform never became measurable and thin, in the sense described above. It remained a deeply personal experience rooted in knowing the other. This is what Paula Bialski wrote about her first hosting experience on CouchSurfing: "He [the guest] would speak, and I would often listen. It was the first time I ever invited a stranger into my home, and the first time I ended up speaking to a stranger until the late hours of the night."<sup>45</sup>

It may be useful to think about the transformation of trust as an instance of rationalization. According to Max Weber, rationalization is the process through which more and more spheres of life become subject to calculation, measurement, and control

---

<sup>43</sup> Established in 2003, it reached 1 million users in 2009 and would go on to sign up more than 10 million users by 2015. See Coca Nithin, *The Improbable Rise and Fall of Couchsurfing*, THE TRAVEL CLUB (June 12, 2015), <https://www.thetravelclub.org/articles/traveloscope/698-the-improbable-rise-and-fall-of-couchsurfing/>. Notwithstanding, the platform was never profitable and became mired in several legal controversies. The platform is now a marginal player in the two-sided marketplace segment. Yet CouchSurfing remains important for its pioneer role in creating a different way to travel.

<sup>44</sup> Patricia Marx, *You're Welcome*, THE NEW YORKER (Apr. 16, 2012), <https://www.newyorker.com/magazine/2012/04/16/youre-welcome>.

<sup>45</sup> PAULA BIALSKI, BECOMING INTIMATELY MOBILE (2012).

(Weber, *Science as a Vocation*).<sup>46</sup> While rationalization brings a leap in efficiency, Weber was deeply ambivalent about the process. Alongside increased efficiency, Weber noted that rationalization ushers a process of disenchantment. Disenchantment refers to the semiotic changes that result from the application of measurement and calculation to actions and situations that were previously less rationalized. Measurement, in other words, does more than reflecting the state of the world. It also changes its meaning. For instance, the measurement of economic activity, like gross domestic product (GDP), had not only rendered a previously abstract entity (the economy) visible and actionable in ways previously unimaginable, it also fundamentally shifted how policymakers think about “the economy.” Once developed and implemented, economic growth became an end of its own, and policymakers now design interventions designed to boost GDP growth.<sup>47</sup>

Similarly, the development and widespread implementation of intelligence tests had altered the meaning of wisdom. From a holistic attribute of a person, appreciable in conversation or through an in-depth acquaintance, being smart is gradually reduced to solving a series of relatively meaningless multiple answer questions at a given speed.<sup>48</sup> Note that in both examples, the development of measurement procedures does little to clarify the terms’ ambiguities. The economy remains an abstract concept whose boundaries are imprecise, and wisdom remains an elusive and confusing attribute. More pointedly, the development of measurement procedures to capture “the economy” or “intelligence” reduced their meaning. From wholesome concepts that resist measurement but convey deep

---

<sup>46</sup> Max Weber, *Science as a Vocation*, 87 DAEDALUS 111 (1958).

<sup>47</sup> Timothy Mitchell, *Economentality: How the Future Entered Government*, 40 CRITICAL INQUIRY 479 (2014).

<sup>48</sup> NIKOLAS S. ROSE, *GOVERNING THE SOUL: THE SHAPING OF THE PRIVATE SELF* (1999).

meaning, these concepts' measurement turned them measurable but almost meaningless.<sup>49</sup>

The measurement and modeling of trust, especially integrating these models into two-sided markets, is affecting a similar transformation. Trust remains an abstract concept, slippery and full of ambiguities. Still, now platform operators can respond in real-time to challenges and mistrust and actively optimize two-sided markets to generate trust. The type of trust that emerges through such intervention is different from the interpersonal trust that actors skillfully develop through repeated interpersonal exchanges. In place of interpersonal reciprocal trust, this new type of trust results from the accrual of reviews of past transactions with third parties. This trust, as we have seen, is mostly impersonal. It is not based on past interaction with the trusted actor but on others' experiences. It is based on the position of the trusted actor relative to other actors on the platform.

Importantly, this trust is also thin because it is domain-specific. This specificity of trust is, in part, merely a function of the organization of reputation systems, which are nested within specific two-sided markets (Uber, AirBnB, Amazon, etc.). But it is also a byproduct of the fact that this type of trust is impersonal and detached from the actors' actual past experiences.

Domain-specific trust is not a new phenomenon. We typically trust our doctor's advice on matters that pertain to health but will be quite cautious when it comes to assessing her stock purchase advice. We trust our lawyer (always a mistake) for legal advice, not health matters. Yet, the domain-specific trust that

---

<sup>49</sup> Researchers sometimes argue that intelligence is precisely what we measure in psychometric test. See Claude S. Fischer et al., *Understanding 'Intelligence'*, in *INEQUALITY BY DESIGN: CRACKING THE BELL CURVE MYTH* 22 (1996).

reputation systems generate does not depend on diplomas or other forms of credentialed expertise. Relatively thin trust relies on the labor of previous reviewers.

Finally, whereas interpersonal trust relies on the parties' skilled interaction, the mediating role that reputation systems play in creating and sustaining this type of trust, means that some of the skill involved in creating the trust does not reside between the parties that exchange goods and services. Instead, this skill is appropriated, in part, by the Trust and Safety teams or other platform operators that continuously experiment and optimize their systems. To the extent that this is the case, thin trust operates "under the hood" or outside the consciousness of the involved actors. The development of reputation systems powers a leap in the scale of trust relations, but it leaves the parties to the exchange without a clearer understanding of the conditions within which they live and act.

#### **TOWARD APPLIED SOCIAL SCIENCE**

This chapter explores the development of two-sided markets, focusing on how technologically sophisticated reputation systems foster the creation of thin trust between actors on those platforms. The chapter also calls attention to a new frontier for the social sciences. In the past, social research was an academic pursuit. To the extent that social scientists found their way to industry, their roles were typically marginal and confined to consumer behavior studies through surveys or focus groups. However, the digitization of everyday life creates entirely new possibilities for the integration of social science and business. Along with these possibilities come new ethical questions and risks. This last section returns to the issue of trust to explore these new possibilities and dangers.

Data exists on things that used to be beyond the reach of quantification and experimentation. The range of questions that

social scientists can now ask has expanded. More importantly now it exists as a mechanism through which interactions can be planned and their consequences measured. Such is the nature of socio-technical systems, and social scientists could be a part of the solutions that get designed.

Social phenomena depend on the interaction of multiple actors. Until recently, it was practically impossible to intervene and experiment within such interactions.<sup>50</sup> The digitization of social life changes allows social scientists to experiment on a very large scale. Importantly, the realism of these interventions, since we operate on the same platforms and interfaces actors use in their everyday life and in the same settings, is very high.

The digitization of social life presents social scientists with an exciting research frontier. More than that, the mediated nature of online platforms effectively allows social scientists not only to study but to intervene and curate social interactions at scale. A digitized social space means a space where operators can plan and measure every interaction. The analogy with urban planning is apt. In this newly digitized space, operators can experiment and optimize interactions in the same way that urban planners design urban spaces and traffic flow, but with far better efficiency.

If trust could be measured and modeled, it can also be manipulated in more invasive ways. For instance, in 2014, Uber launched a carpooling service on its app, allowing users to share a ride. However, putting strangers in the same car is a tricky socio-technical feat. Part of the challenge was to match riders to correct routes efficiently. But Uber quickly discovered that bad matching of

---

<sup>50</sup> Social psychologists attempted to do that by treating the individual as the locus of the experiment, but the *raison d'être* of the social sciences is to study relations between individuals.

riders could result in unpleasant altercations and unexpected challenges to drivers.<sup>51</sup>

Friction in the interface between service providers is nothing new, but in the past, companies had limited ability to respond to these breaches of trust in real-time. Uber or its likes could have intensified background checks and ban problematic drivers or passengers, and it could introduce new rules of behavior in rides. But powered with a good trust model, platform operators could have introduced a whole new roster of interventions to prevent the problem. With better modeling of drivers and consumers, Uber now could have prevented matchings of incompatible riders or drivers, it could have identified difficult times or areas of service, and it could have created changes in the app itself to help consumers report challenging encounters in real-time (which could be used to further optimize the model). None of these possibilities existed before, and for sure, Uber did not deploy the solutions we cursory mentioned above. Yet, they remained possible and platform operators could test whether any of these solutions worked, and to do that in an extremely short interval. A lack of trained social scientists is the main reason why these solutions were not tested.

Uber and similar platforms operating in two-sided markets are modeling interactions. The consequences of their products do not remain confined within the virtual worlds they create. Instead, their products intervene and alter people's social interactions. Two-sided markets have created opportunities for social scientists to measure and design social interactions at a scale not previously possible. While exciting, these developments pose challenging ethical questions. Using trust as an example, users identified as a

---

<sup>51</sup> Kiana Cornish, 'Ride from Hell': Carpooling in the Age of Uber Can Be...Awkward, WALL ST. J (Dec. 6, 2018), <https://www.wsj.com/articles/ride-from-hell-carpooling-in-the-age-of-uber-can-beawkward-1544112559>.

low trust based on their actions on the platform could be exposed to different conditions to win back some of their trust. Yet, the model upon which trust levels are predicted ignores the personal reasons why users may have different levels of trust. Intervening to bypass barriers to trust may become a manipulation that reduces individual choices. Informing users about the potential existence of such models is only a first step in protecting users' freedom.

A better and more systematic approach to address ethical questions would require the platforms to leverage social scientists' expertise in designing and planning products like a reputation system. Social scientists are uniquely capable of understanding the impact of socio- technological systems. For example, when Nextdoor was trying to find a solution to racist comments on its platform, they hired Jennifer Eberhardt, a Stanford social psychologist. Nextdoor CEO described Eberhardt's work:

The basis of her research is around something she calls decision points. If you make people stop and think before they act, they probably won't do the racist things that they do." Today, if you post in the crime section and decide to use race to describe a person, the platform makes you fill in two other characteristics. This simple intervention reduced racist posts by 25% in 2016.<sup>52</sup>

However, social scientists were not included in designing the app from its beginning; neither were they part of its measuring and monitoring. Instead, Nextdoor stumbled upon the solution after other approaches failed and the community faced significant strife. The penetration of technology into more life domains has created the space for applied social science.

---

<sup>52</sup> Pendarvis Harshaw, *Nextdoor, the social network for neighbors, is becoming a home for racial profiling*, SPLINTER (Mar. 24, 2015, 10:02 AM), <https://splinternews.com/nextdoor-the-social-network-for-neighbors-is-becoming-1793846596>.