Fostering trust: when the rhetoric of sharing can backfire

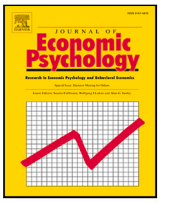
Cicognani, S.; Romagnoli, G.; Soraperra, I.

Journal of Economic Psychology 102 (2024) 102728

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep



Fostering trust: When the rhetoric of sharing can backfire

Simona Cicognani [a,*], Giorgia Romagnoli [b], Ivan Soraperra [b]

[a] Department of Economics, Leiden University, The Netherlands
[b] Amsterdam School of Economics - University of Amsterdam, The Netherlands

ARTICLE INFO

Dataset link: https://osf.io/8udhp/?view_only=

JEL classification:
C92
D84
D91

Keywords:
Trust game
Sharing economy
Framing
Communication
Experiment

ABSTRACT

The peer-to-peer lending and transfer of underutilized resources – the so-called Sharing Economy – make up for a sizeable and rapidly increasing portion of the economy. The digital platforms that enable it have the primary goal of promoting trust among users and providers. In an online experiment, we study how the platform's revenue model (*No-profit* vs. *For-profit*) interacts with its communication strategy in shaping trust. In particular, we consider two communication strategies, one highlighting the trust dimension and the other highlighting the interaction's profitability. We show how these strategies can affect trust differently for pure-sharing and market-oriented for-profit platforms. We observe that a communication strategy that evokes trust feelings, if sent by a for-profit platform, decreases trust in the interaction. This evidence suggests that leveraging a rhetoric of trust can backfire for profit-oriented platforms in terms of trust generation.

1. Introduction

The past two decades have seen a rapid growth of the peer-to-peer transfer of underutilized resources, the so-called Sharing Economy. While the sharing model *per se* is thousands of years old, it has seen a surge in modern times thanks to the growing popularity of the internet and is now serving a sizeable portion of the demand in sectors as diverse as the automotive, hospitality, retail, consumer goods, and entertainment industries. The Sharing Economy was worth 15 billion U.S. dollars in 2014 and is projected to grow to 335 billion U.S. dollars by 2025 (Vaughan & Hawksworth, 2014). The common denominator to virtually all modern sharing-economy ventures is the existence of (digital) platforms that operate as connectors between providers and users. These platforms follow remarkably diverse structures and business models. Some of these, such as Couchsurfing at its onset, are so-called pure-sharing platforms and do not extract profits. In contrast, most others, such as Airbnb, Uber (and Couchsurfing itself after 2011), are market-oriented for-profit platforms often generating billions of dollars in yearly revenues. For instance, the primary source of Airbnb's revenue stems from service fees from bookings charged to both guests (usually under 14.2%) and hosts (3% for every completed booking), while Uber makes profits by charging a commission on the driver's fare (about 25% of the fare or more).

Invariably, the primary function of these platforms is to increase the level of trust between users and providers (Hofmann et al., 2022; Ter Huurne, Ronteltap, Corten, & Buskens, 2017), so much so that the Sharing Economy has been alternatively labeled the Trust Economy (PWC, 2015). To achieve this goal, most Sharing Economy platforms have tapped heavily into the rhetoric of sharing, access, community, and trust, concepts that are loaded with positive and even anti-capitalistic connotations. For instance, on Uber's webpage, one can find the following sentences: "Good things happen when people can move, whether across towns or toward their dreams. [...] What started as a way to tap a button to get a ride has led to billions of moments of human connection [...]".. Adopting this language has met some criticism for its alleged misplacement vis-à-vis earning models which are often quite traditional and

* Correspondence to: Steenschuur 25, 2311 ES Leiden, The Netherlands.
E-mail addresses: s.cicognani@law.leidenuniv.nl (S. Cicognani), g.romagnoli@uva.nl (G. Romagnoli), i.soraperra@uva.nl (I. Soraperra).

https://doi.org/10.1016/j.joep.2024.102728
Received 20 February 2023; Received in revised form 5 March 2024; Accepted 9 March 2024
Available online 16 March 2024

very lucrative for the platform owners (Codagnone & Martens, 2016; Eckhardt & Bardhi, 2015; Jemielniak & Przegalinska, 2020). The way the platform positions itself and its communication strategy are essential aspects that can have a decisive impact on trust. Moreover, the same communication strategy can have a different effect depending on the revenue model of the platform and, precisely, depending on whether the platform is pure-sharing or for-profit. According to the marketing and management literature, non-market strategies are extremely relevant for digital trust building, especially in contexts characterized by underdeveloped formal regulations (Chen & Wang, 2019; Cheng, Fu, Sun, Bilgihan, & Okumus, 2019; Ko, Amankwah-Amoah, Appiah, & Larimo, 2022).

In this study, we experimentally investigate how the earning model of the platform interacts with its communication strategy in shaping trust. We build on the experimental notion of trust, augmenting a standard trust game (Berg, Dickhaut, & McCabe, 1995) by introducing a third player, which aims at proxying the sharing platform. We exogenously manipulate two dimensions: (i) the communication strategy of the platform and (ii) its monetary incentives.

The first treatment variation, i.e., the communication strategy of the platform, refers to the pre-formulated message that the third player can choose to send or not to send. In the *Money* treatment, the message stresses the higher monetary rewards that can be gained with a high level of trust. In the *Trust* treatment, the focus of the message is diverted away from the material domain and directed to evoke positive feelings of trust and reciprocity. A third treatment (called *Baseline*) does not allow the third player to send any message to evaluate the effect of communication separately.

The second treatment variation concerns the monetary incentives of the third player. In the *No-profit* condition (intended to replicate the pure-sharing platforms), the third player earns a flat fee unaffected by the amount of trust displayed in the trust game. In the *For-profit* condition (intended to replicate the for-profit platforms), the third player earns a fee proportional to the amount of trust displayed in the trust game. We thus have six between-subject treatments based on the interaction between the type of communication (*Baseline* or No message, *Trust* message, *Money* message) and the revenue model (*No-profit*, *For-profit*).

Our goal is to examine whether the two communication strategies affect trust generation differently in pure-sharing and market-oriented for-profit platforms. This is a relevant topic given the recent debate on true sharing or pseudo sharing (Belk, 2014a, 2014b; Botsman, 2013; Eckhardt & Bardhi, 2015), centered around the issue of whether or not some lucrative platforms can be considered as part of the Sharing Economy (Phua, 2019; Schor & Vallas, 2021). Since many users on sharing platforms are guided by anti-capitalist and anti-consumption sentiments (Hamari, Sjöklint, & Ukkonen, 2016), indulging in the rhetoric of trust while the platform is based on pure capitalist principles can be detrimental to trust generation. This represents our main conjecture, and it is inspired by recent discontentment by users and activists around this topic, for instance, after Couchsurfing shifted from a pure-sharing activity to a for-profit one (The Guardian, 2011).

Our results show that trust is more prominent in the presence of communication. More importantly, we find that while in the *No-profit* condition, the type of message is inconsequential for the level of trust, in the *For-profit* condition, the *Trust* message achieves significantly lower levels of trust than the *Money* message. These results suggest that leveraging a trust rhetoric can backfire for profit-oriented platforms. However, we should use a word of caution here in that we find significant results for the simple effects but not when testing for an interaction effect. The non-significant interaction may be ascribed to a potential lack of power when using the more demanding Diff-in-Diff approach (see more details in Section 5.2 and Table 2). Future research is needed to evaluate the robustness of the result to different specifications.

In the second part of the paper, we focus on potential reasons behind this pattern of results with a new pool of subjects and exclusive focus on the *For-profit* conditions. More specifically, we are interested in examining three non-mutually exclusive mechanisms that could explain our main results. First, we explore whether people find the rhetoric of sharing to hide traditional corporate objectives of profit maximization to be misused and insincere. Second, we explore whether the communication strategy of the platform affects participants' higher-order beliefs regarding the intentions and trustworthiness of their fellow platform participants. Lastly, we investigate whether there exists a desire to punish the platform for adopting a communication strategy that uses the rhetoric of sharing to make profits. We find that the lower level of trust associated with the *Trust* message cannot be explained by trustors expecting a lower level of reciprocation from the trust receivers. Instead, we find some evidence of the *Trust* message being perceived as less accurate and honest than the *Money* message.

We extend the existing literature in three main ways. First, by running a lab experiment to model a Sharing Economy context, we can investigate what matters for trust generation in a controlled setting. This constitutes an improvement compared to simulation studies in that we test our behavioral conjectures on real subjects and across randomized treatments. Second, we introduce a third party in the trust game. The third party has financial stakes in the interaction between the trustor and the trustee and is distinguished by a given communication strategy that varies across treatments. This way, we provide the first study able to pinpoint the effect on trust of the interaction between the type of communication and the payoff structure. Third, we add to the literature of three-person experimental trust games that consider trust decisions involving multiple agents (see Section 2 for further details).

The paper is structured as follows. Section 2 reviews the relevant literature; Sections 3-5 cover the first experiment; the second experiment is presented in Section 6, and Section 7 discusses and concludes.

2. Related literature

This section first illustrates the concept of trust used in the sharing economy and in the experimental economics literature. It also reviews the experimental studies that are most related to our work and depicts how we depart from them.

The literature on the sharing economy has initially focused on studying trust in business-to-consumer (B2C) e-commerce (McKnight & Chervany, 2001), and lately started to investigate the role of trust in consumer-to-consumer (C2C) e-commerce (Leonard, 2012). Recent contributions differentiate between three different targets of trust: peer, platform, and product (Hawlitschek, Teubner,

& Weinhardt, 2016). Trust in the context of the sharing economy is indeed a multi-faceted and multi-dimensional concept. According to Hawlitschek et al. (2016), the dimensions of trust relate to ability, integrity, and benevolence. The latter two dimensions are highly relevant in C2C interactions for two main reasons: (i) regulative buyer protection is still not widespread (Koopman, Mitchell, & Thierer, 2014) and (ii) customers in typical C2C interactions are in vulnerable positions (e.g., accommodation or car sharing). The sharing economy's key feature is a platform that acts as a mediator between buyers and sellers (the peers). Trust in the platform itself could also boost the interaction between peers. For instance, as pointed out by Hawlitschek et al. (2016), the fact that the platform handles data reliably and securely, how much it charges the users, the extent of user support, and its reputation are valuable factors worth considering for the buyers and sellers.

This notion of trust, widely used in the sharing economy literature, is related to the one employed in behavioral economics and specifically modeled with the trust game (Berg et al., 1995). Here, a trustor voluntarily places resources at the disposal of the trustee without any legal commitment from the latter, thus making themselves vulnerable to the trustee's actions (Bicchieri, Duffy, & Tolle, 2004; Hong & Bohnet, 2007). As such, trust is related at least to some extent to the expectation that trusting the trustee will pay off financially (Coleman, 1990; Fehr, 2009). The trustee indeed can reciprocate the trust received by returning a large sum to the trustor at a personal cost. This represents well the incentives faced by the providers of goods and services in the sharing economy: they often can deliver a better service only through higher investments of time and money, thus at a material cost to themselves. The sharing platform can promote this trust dynamic with various tools, such as information provision, customer protection, and communication strategy. The last dimension is the focus of our paper. We thus believe that the trust game, augmented with a third player that decides the communication strategy, encapsulates the essential ingredients to explore the effect of communication on the dynamic of trust in the sharing economy. At the same time, its simplicity and stylized nature offer a complementary methodology to study trust in the sharing economy, allowing us to isolate the effect of communication while holding all other aspects unchanged in a controlled environment.

In the recent literature on the trust game, several papers have investigated trust to address particular issues departing from the standard two-person environment (Banuri's, de Oliveira, & Eckel, 2019; Bigoni, Bortolotti, Casari, & Gambetta, 2013; Cassar & Rigdon, 2011; Rietz, Sheremeta, Shields, & Smith, 2013; Sheremeta & Zhang, 2014). For instance, Cassar and Rigdon (2011) create a game with one trustor and two trustees, and another one with two trustors and one trustee, to examine peer effects in trust and trustworthiness.[1] Conversely, Sheremeta and Zhang (2014) and Rietz et al. (2013) add to the standard dyadic trust game structure an intermediary to resemble a scenario with an investor, an intermediary, and a borrower. This dynamic structure is also present in Banuri's et al. (2019), which consider a modified trust game in which a principal can transfer resources to an agent, who can then transfer resources to a needy recipient in an attempt to proxy a care work setting. To the best of our knowledge, no study has looked at the consequences of introducing a third player in the trust game who can send messages or suggestions, which represents our contribution to this stream of literature. As in Banuri's et al. (2019), our modified trust game aims to represent an applied setting (i.e., a sharing economy context) in which trusting someone entails making yourself vulnerable.

There is a well-established literature documenting that decisions to trust strangers may not depend only on economic dynamics, such as expectations that trust will be honored and the size of the reward for trusting, but also on emotional and social ones (Dunning, Fetchenhauer, & Schlösser, 2012). According to Bicchieri et al. (2004), every time we decide to trust someone, we pose ourselves in a social exchange involving uncertainty and vulnerability. Therefore, the trust game has both a strategic component, connected to risk preferences and the belief about other people's trustworthiness, and a more psychological one, connected to social preferences of various forms, including fairness, reciprocity, altruism, and positional concerns (Wuthisatian, Pingle, Nichols, et al., 2017). In a seminal article that builds on neurobiological and experimental evidence, Fehr (2009) shows that merely the strategic component cannot capture trust and that social preferences play a major role.

Framing and communication are factors that can affect the psychological component of trust, as attested by a large literature discussed in the next paragraph. Normally, in this literature, the framing is chosen either by the experimenter or the players directly engaged in the trust game. The distinctive feature of our paper is introducing a third party with financial stakes in the game that chooses the framing, i.e., the communication strategy. While this is the appropriate choice given our intention to model the sharing economy (where the style of communication that frames the transaction is largely chosen by the platform), this feature also represents a novelty in the trust game literature.

Deutsch (1958) pioneered the study of framing effects on trust. Using a series of two-person games similar to a prisoners' dilemma (PD), the author studies the effect of "motivational orientation"—i.e., giving different motivations to the subjects' choice in the game—on trust. Results show that when actions are motivated with cooperative rhetoric, reciprocal trust (cooperation) increases significantly compared to those motivated using individualistic or competitive rhetoric.

Cronk (2007) was the first to study rhetoric framing in the trust game. Participants are Maasai, and the game is framed neutrally or as the "Osotua" game. In the Maasai culture, Osotua is a relationship or bond in which two partners can ask for help and commit to providing help when needed. The paper does not find significant differences even though the Osotua framing reduces the amount sent and the amount sent back. Chaudhuri, Li, and Paichayontvijit (2016) use a standard trust game and manipulate the protocol and instructions to understand whether "goal framing", i.e., clarifying the consequences of the actions and what they entail, affects trust. The "goal framing" treatment "clearly articulates the conflict between individual self-interest and the social optimum" (Chaudhuri et al., 2016, p. 129) and is associated with a higher level of trust. Explicitly using the words "trust" and

[1] Also Bigoni et al. (2013) employ an expanded trust game form, with one trustor and two trustees, that they dub Collective Trust Game.

"reciprocity" (i.e., context-loaded treatments), hence resorting to emotive words, marginally increases the amount compared to context-neutral instructions.

Liberman, Samuels, and Ross (2004) study social framing in repeated prisoner's dilemma games and prove that social framing matters. The cooperation rate increases substantially when the game is labeled the *Community game* instead of the *Wall Street game*. In a subsequent study using a one-shot prisoner's dilemma game, Ellingsen, Johannesson, Mollerstrom, and Munkhammar (2012) show that this is not because the framing changes social preferences or the social image of the two actions but because it changes beliefs when preferences are intention-based. Specifically, frames help coordinate cooperation. In a follow-up paper, the same authors show that the effect is mostly driven by females who seem more sensitive to social framing (Ellingsen, Johannesson, Mollerstrom, & Munkhammar, 2013). Finally, Chang, Chen, and Krupka (2019) showcase the role of framing in a dictator game setting. Specifically, the authors show that Democrats and Republicans have a different willingness to redistribute only when the choice is framed as a tax decision. The difference disappears when the choice is neutrally framed as an allocation decision. The authors argue that the tax frame triggers the social identity of the participant and, consequently, the corresponding decision.

Our paper also relates to the literature on advice and communication in trust games. Deutsch (1958, 1960) provides evidence of the positive effect of communication on trust. In a series of treatments, players can send prepared notes to the other players. These notes focus on trust and retaliation, and they help increase trust compared to a baseline setting with no communication. In contrast, our study focuses on a setting where different forms of communication strategies are compared to each other. Moreover, we manipulate both the type of communication (*Trust* vs. *Money* message) and the type of (lucrative) context in which the message is sent (*For-profit* vs. *No-profit*) in an attempt to disentangle better the dynamics of trust generation. Also, Capraro, Vanzo, and Cabrales (2021) find evidence that framing matters in pro-social decisions but in their setting, they let players in different types of dictator games choose descriptions that may be potentially beneficial for themselves. This beneficial component is closely connected to our setting, although we provided pre-formulated messages and let players choose between not sending a message and sending a message that could benefit them.

Within the communication literature in experimental games, our paper is also related to the literature on compliance with a recommendation or request made by an authority. Significant experimental evidence suggests that people pay taxes to obey an authority. In a tax evasion experiment, Cadsby, Maynes, and Trivedi (2006) compare tax compliance in an obedience-to-authority setting and in an invitation-to-gamble setting, and they find a higher degree of compliance in the former than in the latter setting. In many settings, the authority is chosen to be the experimenter, who provides nudges to subjects making decisions over different scenarios (Karakostas & Zizzo, 2016; Pelligra, Reggiani, & Zizzo, 2020; Sonntag & Zizzo, 2015). We depart from this in that in our setting, the role of authority sending recommendations is taken by the third player, who, in our intentions, is a proxy for the sharing economy platform.

We also contribute to recent literature that proposes leveraging on social dilemmas to model online markets and sharing economy scenarios (Corten, 2019; Tadelis, 2016). Chica, Chiong, Adam, and Teubner (2019) propose an agent-based model incorporating an evolutionary trust game that manipulates providers' and consumers' trustworthiness and adds to the picture the possibility of imposing a penalty on untrustworthy providers and providing insurance for consumers. In contrast to these papers, which theoretically propose applying the trust game or applying it in a simulation setting, we perform an experiment with real subjects making decisions in randomized controlled treatments.

3. Experimental design and predictions

3.1. Experimental design

The design of the experiment models the interaction between users and providers as a stylized trust game with a sender, player A, and a receiver, player B (Berg et al., 1995) (for a thorough meta-analysis on trust games, see Johnson & Mislin, 2011). The sender is endowed with 100 points and decides how many of the points s/he wants to send to the receiver. The sender can send none, part, or all of his/her points to the receiver. The points sent to the receiver are multiplied by three so that the receiver obtains three times the amount sent by the sender. After the sender makes her decision, the receiver can send back none, part, or all of the points received. The amount sent by the sender is an accepted measure of how much the sender trusts the fact that the receiver will send back some money.

We modify this classic game by introducing a third player, player C, also referred to as the platform, for exposition. Before the sender and the receiver make their decisions, the platform can communicate with them. Specifically, the platform can send a (pre-formulated) message to both the sender and the receiver to suggest the benefits of sending and returning points. If the platform decides to send a message, this is delivered to 20 different couples of senders and receivers.

We evaluated the pros and cons of adopting pre-formulated messages. Pre-formulated messages allow the experimenter to have more control over the type of message players Bs see. The advantage of having more control is that each player B is subject to the same information; therefore, if we observe differences in behavior in players Bs, these are due to the different kinds of messages they are exposed to. As noted by Brandts, Cooper, and Rott (2019), pre-formulation is to be preferred to free-form when: (i) a specific aspect of communication is of interest (in our case, the trust component of the message), and (ii) the experimenter wants to control the communication content. However, the advantage of a higher control comes at the cost of diluted effects in observed behavior, as pointed out by Charness and Dufwenberg (2010) and Lundquist, Ellingsen, Gribbe, and Johannesson (2009). For instance, Charness and Dufwenberg (2010) report that pre-formulated promises, which they call "bare promises", have a lower effect on behavior than free-form messages.

The experiment follows a 3 × 2 design where we manipulate two dimensions: (i) the type of message player C can send to A and B and (ii) the payoff structure. As for the message, we have three treatments: *Baseline*, *Trust*, and *Money*. In the *Baseline* treatment, player C cannot send messages. In the *Trust* treatment, player C can send the following predetermined message to A and B:

> "Hi guys. I think you should do the following: participant A should send all the points as this will show how much he trusts participant B; participant B should return 50% of the points as this will show his gratitude to participant A".

Instead, in the *Money* treatment, player C can send the following predetermined message to A and B:

> "Hi guys. I think you should do the following: participant A should send all the points, and participant B should return 50% of the points. This makes you earn more money".

The two messages aim to capture two different rhetorical strategies the platform uses. Importantly, senders and receivers are not informed about the possibility available to player C. They are simply told that C has the opportunity to send a message and that the message sent is delivered to 20 couples of A and B.[2] Therefore, they do not know that player C has scripts available as a set of options. This script feature does not interfere with our primary focus: studying the effect of rhetorical messages on trust.[3]

The payoff structure follows one of two settings: one capturing a pure-sharing no-profit platform and the other capturing a market-oriented for-profit platform. In the *No-profit* setting, player C is paid a fixed amount of points independently of the amount sent by A. Specifically, C obtains 10 points for each of the 20 couples. This amounts to a fixed payoff of 200 points. In the *For-profit* setting, player C keeps 10% of the points passed to the receivers for each of the 20 couples. So, for instance, if A sends 60 points to B, these are tripled, resulting in 180 points. So, player C will keep 18 of the 180 points, and the remaining 162 will be given to B. In the *For-profit* condition, player C collects the points independently of whether he sends a message or not and independently of the choice of player B. Importantly, players A and B are fully informed about the payoff structure for player C in all treatments.

We decided to take money off player A and give it to C in the *Profit* conditions for realism concerns: most Sharing Economy platforms make money by charging the parties a fee, thus effectively reducing the gains of trade that are left to the parties to share. Removing this feature would have changed the nature of the game in a way that could have affected the main measures in our experiment: it is possible that players A and B would have a better reaction to a for-profit platform using the rhetoric of trust if this platform were not profiting directly from them.

Regarding theoretical considerations of the game we implement, the game at the basis of our setting is a standard trust game. Theoretical predictions under the classic assumption of selfishness and rationality are the usual. If, instead, we assume outcome-based other-regarding concerns, we should observe the same trust level no matter the message. Here, we assume that preferences over outcomes do not change with the message and that the sender has correct beliefs about such preferences. If we observe any effect, this means that the message either has an effect on the trustor's preferences (who may feel some more warm glow) or has an effect on his/her beliefs about the amount returned by the trustee.

3.2. Behavioral predictions

We cannot directly compare trust across the two different payoff structures in terms of analysis. This is due to the fact that for a fixed choice of A and B, the payoff distribution and the payoff received by all players differ in the two settings. This may impact the beliefs and the decision of the sender. This, however, is not the objective of the experiment. The main aim is to compare the effect of different communication strategies on trust while keeping the type of platform constant, i.e., the payoff structure.

From this perspective, we formulate the following hypotheses:

Hypothesis 1a. When the platform is *for-profit*, using a communication strategy based on trust can backfire compared to using a communication strategy that focuses on reciprocal gains. In other words, we expect a lower amount sent by A towards B in the *For-profit*&*Trust* condition.

Essentially, we conjecture that leveraging a rhetoric of trust when the platform benefits monetarily from it may trigger negative feelings about the platform's intentions.

Hypothesis 1b. When the platform is *no-profit*, we expect no difference when the communication strategy is based on trust compared to when the communication strategy highlights the reciprocal gains.

This second hypothesis acts as a background check that the differential effect of the two types of messages is not an absolute feature but, instead, depends on the earning model of the platform.

[2] This may have changed the interpretation of silence: for example, player C may have chosen silence because she did not like the message, but this inference would be lost for the receivers of the message.

[3] The reason to let player C not send a message is two-fold. First, to avoid deception: we felt that saying player C has "sent you a message" (when player C was, in fact, forced to send the message) would not be truthful. Second, for realism: real-world platforms are not forced to build a distinctive communication language. For example, many second-hand websites (considered part of the sharing economy) adopt a very plain interface that does not promote either the rhetoric of trust or the rhetoric of money.

4. Participants and procedures

We programmed the experiment in oTree (Chen, Schonger, & Wickens, 2016) and then conducted it via the online labor market platform MTurk (for more details on MTurk, see Horton, Rand, & Zeckhauser, 2011). Besides the similarity of results of experiments conducted on standard subjects (university students in physical labs) and MTurk workers (Fréchette, Sarnoff, & Yariv, 2022; Hauser, Paolacci, & Chandler, 2019), our setting depicting a sharing economy context is particularly suited for an online set-up. To check the understanding of instructions by participants, which is particularly relevant in one-shot games, we run a set of five control multiple-choice questions, each one reporting three alternatives (we report the instructions and the control questions in the Online Appendix).

In order to be eligible, participants had to meet the following requirements: (i) they had to have completed at least 100 HITs; (ii) they had to be registered with an American account; and (iii) they had to have an approval rate of at least 85%.[4]

We recruited a total of n = 1238 workers, assigned to one of the six treatments (between-subjects design), ending up with 615 workers in the 3 *For-profit* conditions (205 in *Trust*, 205 in *Money*, and 205 in *Baseline*) and 613 workers in the 3 *No-profit* conditions (202 in *Trust*, 207 in *Money* and 202 in *Baseline*).[5] Overall, we ran 22 sessions (between July 2018 and July 2019), with 2 sessions in the *No-profit Baseline* and in the *No-profit Money* treatments, 4 sessions in the *Profit Money* treatment, 5 sessions in the *Profit Trust* treatment and 3 sessions in the remaining treatments. Each session was on a different hit, so subjects were assigned to treatments according to the hit they clicked on. The first and second sessions in each treatment aimed at 80 to 100 subjects (depending on the choice of player C).[6]

Since data collection depended on when the workers accepted the hit and on the dropout during the experiment, the subsequent sessions varied in size depending on how many participants were still needed to complete the data collection. In order to be matched, participants had to complete the control questions. After that, they were put in a waiting room. Participants entering the waiting room answered all control questions correctly. If a participant was already waiting, the two participants were matched and started the game. If not, the participant had to wait up to 5 min. If no one was showing up, the participant was sent directly to the second stage. Of the participants taking part in each condition, five played the role of player C, and the remaining played the role of player A and player B. This means that we collected data from about 100 couples per condition.

The experiment included two parts. In the first part, participants played the trust game described above. They were endowed with 100 points in the second part and played a dictator game with a charity of their choice (the available options for the charity donation were WWF, Doctors without Borders, and UNICEF). The rationale for the dictator game is twofold: (i) to introduce a control measure for the level of altruism of participants; (ii) to provide them with a game to play in the scenario in which they did not find a partner in the trust game who had already completed the control questions. Since the latter reason is the driving factor of the decision to include a second individual task, we opted for having a charity instead of another participant playing the game.

A short demographic questionnaire followed these two parts (56% of the sample is represented by males, the average age is 35 years old, the median education is undergraduate studies, and 90% of the sample has a job). At the beginning of the experiment, we informed participants that there would be two independent parts to earn a payment, but we did not inform them of their content.

The experiment took about 15 min to complete. Participants were paid a fixed fee of $0.50 and could earn a bonus that depended on their decisions in the two parts. Specifically, the bonus was determined by converting the total number of points accumulated in the two tasks at the rate of $0.50 for 100 points. The average earnings were about $1.40.

5. Results

In what follows, we will focus on the effects of the messages on trust from a within-condition perspective. Unfortunately, due to efficiency concerns, the comparison between the *For-profit* and the *No-profit* treatments is hard to conduct. The results of the sensitivity analysis we conducted for Experiment 1 are reported in the Online Appendix. The data and the analysis scripts are available on OSF.

5.1. Propensity to send the message (player C)

We start with a brief description of player C's choices when s/he had the opportunity to send a message. Overall, there are 20 such players equally divided in the four conditions. All 10 players in the *For-profit* condition decided to send the message (5/5 in *Trust* and 5/5 in *Money*). Of the 10 players in the *No-profit* condition, 8 decided to send the message, 4/5 players in *Trust* and 4/5 in *Money*.[7] The limited number of subjects in the role of player C does not allow us to draw statistical comparisons across conditions. However, this is not the core of our study. In what follows, we will focus on the behavior of A and B subjects.

[4] In the MTurk jargon, a HIT (Human Intelligence Task) refers to "a question that needs an answer: a single, self-contained, virtual task that a Worker can work on, submit an answer, and collect a reward for completing"., whereas approval rates represent the percentage of successful completion of HITs across requesters. Therefore, an approval rate of 90% denotes a worker that 90% of the time s/he initiated a task successfully completed it.

[5] Note that, due to an assignment mistake, a trustor–trustee couple received the *Money* message instead of the *Trust* message. Moreover, we collected 99 instead of 100 couples in the *Baseline* treatment.

[6] To prevent MTurk workers from participating in multiple sessions of the experiment, we included a participation restriction based on the MTurk worker ID.

[7] Only 2 in 20 C players decided to remain silent, and none of them were in the *For-profit* condition. Deciding not to send a message seemed thus related to the absence of stakes for player C in the *No-profit* condition and did not interact with the accuracy of the message.

Table 1
Trust decisions by A by treatment and message type.

Treatment	Message	(1) Amount sent to B	(2) Proportion of A sending	(3) Amount sent to B conditional on sending	(4) n
For-profit	Baseline (No. present)	44.00 (3.27)	0.84 (0.04)	52.38 (3.15)	100
For-profit	Money	65.74 (3.50)	0.88 (0.03)	74.70 (2.85)	100
For-profit	Not sent	–	–	–	–
For-profit	Trust	55.95 (4.07)	0.76 (0.04)	73.62 (3.38)	100
No-profit	Baseline (No. present)	50.76 (3.58)	0.83 (0.04)	61.28 (3.28)	99
No-profit	Money	58.62 (4.37)	0.83 (0.04)	70.87 (3.85)	81
No-profit	Not sent	49.45 (6.01)	0.78 (0.07)	63.81 (5.47)	40
No-profit	Trust	56.04 (4.21)	0.81 (0.04)	69.17 (3.56)	79

Notes: The table reports summary statistics on trusting behavior by As. Column (1) shows the average amount sent to B (means with standard errors in parenthesis); Column (2) shows the fraction of As sending a positive amount, i.e., the extensive margin (proportion with standard errors in parentheses); Column (3) shows the average amount sent by the As that decided to send a positive amount, i.e., the intensive margin (means with standard errors in parenthesis); Column (4) shows the number of participants A in each treatment.

5.2. Analysis of trust displayed by the sender (player A)

We now turn to explore the sender's behavior, or player A. Descriptive statistics of the sender, in terms of average amount sent by treatment and message type, proportion of As sending a positive amount to B, and average amount sent by A by treatment and message type conditional on sending, are reported in columns (1) to (3) of Table 1.

Fig. 1, panel (a) reports the average amount sent by player A in the *For-profit* condition. Here, we notice that sending a message, irrespective of its content, statistically increases the amount contributed by player A. Indeed, a Welch t-test comparing the amount sent when the message is not present, i.e., in the *Baseline*, and the amount sent when receiving a message rejects the null hypothesis of no difference for both the *Money* ($t_{(197)} = -4.537$, $p < 0.001$) and the *Trust* ($t_{(189)} = -2.287$, $p = 0.023$) message. While the *Money* message seems to increase trust more than the *Trust* message, the difference is not significant according to a two-sided Welch t-test ($t_{(189)} = 1.823$, $p = 0.070$). A one-sided test following the direction of Hypothesis 1a, however, rejects the null hypothesis that the amount sent is the same in favor of the alternative hypothesis that the amount sent is higher in the *Trust* compared to the *Money* message ($p = 0.035$).

In panel (b), we split the overall effect of the message on trust into two components: the *extensive* and the *intensive* margin. The former — i.e., the fraction of participants sending a positive amount — is represented by the blue wedge of the pie chart; the latter — i.e., the average amount sent by these participants — is represented by the diameter of the pie. This figure highlights how the difference between the *Trust* and the *Money* message is uniquely visible on the extensive margin. Indeed, while we observe the same fraction of subjects sending a positive amount in *Baseline* and *Money* ($\chi^2_{(1)} = 0.664$, $p = 0.415$), we have a statistically lower fraction of individuals sending a positive amount in the *Trust* treatment compared to the *Money* treatment ($\chi^2_{(1)} = 4.878$, $p = 0.027$). Moreover, conditional on sending a positive amount, contributions are indistinguishable across the two message conditions (Welch two-sided t-test: $t_{(153)} = 0.245$, $p = 0.807$). The findings show that:

Result 1. *In the For-profit conditions, subjects send a higher amount after receiving the Trust message than after receiving the Money message. Moreover, a lower frequency of subjects sends a positive amount after receiving the Trust message than after receiving the Money message.*

We interpret Result 1, which confirms Hypothesis 1a, as the sign that a portion of subjects may be turned off from sending any amount of money after receiving the *Trust* message from a *For-profit* platform, therefore showing a distrusting behavior: subjects might feel betrayed by receiving a *Trust* message given that the platform is lucrative. At the same time, the comparison with *Baseline* shows that sending a message improves trust compared to the case where messages cannot be sent. This would suggest that picking the wrong rhetoric is still better than no rhetoric at all. An important remark here is that the messages also contain suggestions for a specific behavior, which is not present in *Baseline*. So, the improvement may not be due to the rhetoric.

In Fig. 2, we present the analogous analysis for the *No-profit* condition. Here, we verify that, as expected, the *Trust* message no longer has deflating effects on the amount sent compared to the *Money* message (Welch two-sided t-tests: *Money* vs *Trust*: $t_{(158)} = 0.425$, $p = 0.671$; *Baseline* vs *Trust*: $t_{(163)} = -0.956$, $p = 0.341$; *Baseline* vs *Money*: $t_{(163)} = -1.392$, $p = 0.166$). Contributions across the two message conditions are indistinguishable on both the extensive and the intensive margin.[8] Thus, we conclude that in the *No-profit* condition, the *Trust* message does not generally induce lower levels of trust compared to the *Money* message. Therefore, we can conclude that Hypothesis 1b does find confirmation in that:

[8] On the extensive margin, chi-squared tests results comparing the fraction of participants sending a positive amount are the following: *Baseline* vs. *Money*: $\chi^2_{(1)} = 0.000$, $p = 0.984$; *Baseline* vs. *Trust*: $\chi^2_{(1)} = 0.098$, $p = 0.754$; *Money* vs. *Trust*: $\chi^2_{(1)} = 0.078$, $p = 0.780$. On the intensive margin, Welch t-tests results comparing the amount sent by these participants are the following: *Baseline* vs. *Money*: $t_{(137)} = -1.896$, $p = 0.060$; *Baseline* vs. *Trust*: $t_{(138)} = -1.631$, $p = 0.105$; *Money* vs. *Trust*: $t_{(129)} = 0.323$, $p = 0.747$.

Table 2
Regressions: amount sent and decision to send.

	Dependent variable:			
	Amount sent *OLS*		Decision to send *Probit*	
	(1)	(2)	(3)	(4)
d(NoMsg)	−11.950*	−11.437*	0.288	0.250
	(5.203)	(5.389)	(0.204)	(0.212)
d(MoneyMsg)	9.790	11.228*	0.469*	0.526*
	(5.203)	(5.395)	(0.213)	(0.225)
d(NoProfit)	0.088	1.328	0.172	0.223
	(5.538)	(5.754)	(0.213)	(0.222)
d(NoMsg)×d(NoProfit)	6.670	6.385	−0.219	−0.162
	(7.608)	(7.860)	(0.300)	(0.313)
d(MoneyMsg)×d(NoProfit)	−7.211	−8.898	−0.404	−0.468
	(7.805)	(8.063)	(0.314)	(0.330)
$Age - \overline{Age}$		0.154		0.012
		(0.144)		(0.006)
d(Male)		−1.970		−0.350*
		(3.266)		(0.137)
d(Master/PhD)		−0.798		0.392*
		(4.344)		(0.181)
d(Bachelor)		−0.642		0.060
		(3.969)		(0.155)
d(work)		−7.096		−0.438
		(5.956)		(0.288)
Constant	55.950***	63.038***	0.706***	1.169***
	(3.679)	(6.964)	(0.137)	(0.315)
Observations	559	539	559	539
R^2	0.034	0.042		
Log Likelihood			−256.855	−238.844

Notes: Models (1) and (2) report the results of OLS regressions with the amount sent by A as the dependent variable. Models (3) and (4) report the results of two probit regressions with a dummy equal to one when A sends a positive amount as the dependent variable. Explanatory variables include: dummy variables for the message treatments (*d(NoMsg)* and *d(MoneyMsg)* with the **Trust** message treatment as the baseline category); a dummy variable for the *No-profit* treatment *d(NoProfit)*. Models (2) and (4) include a series of control variables: the Age of the participant (demeaned), the gender dummy (*d(Male)*), two dummies capturing the education level (*d(Bachelor)* and *d(Master/PhD)* with "high school or lower" as the baseline category), and a dummy capturing whether the participant is currently employed (*d(work)*).
Sign. codes: *** ≤ 0.001; ** ≤ 0.01; * ≤ 0.05.
Wald tests for the difference between the **Money** and the *Trust* message in the *No-profit* condition: H_0 : 9.790 + (−7.211) = 0 in Model (1) ($F(1,553) = 0.197; p = 0.658$) and H_0 : 11.228 + (−8.898) = 0 in Model (2) ($F(1,528) = 0.152; p = 0.696$).

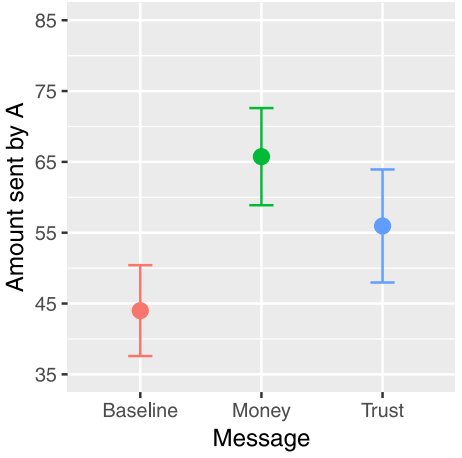
Table 3
Amount sent back to A as a fraction of the amount received by treatment and message type.

Treatment	Message	(1) Amount sent back to A (fraction of amount received)	(2) n
For-profit	Baseline (No. present)	0.27 (0.03)	84
For-profit	Money	0.32 (0.03)	88
For-profit	Not sent	–	–
For-profit	Trust	0.33 (0.03)	76
No-profit	Baseline (No. present)	0.28 (0.03)	82
No-profit	Money	0.38 (0.03)	67
No-profit	Not sent	0.37 (0.05)	31
No-profit	Trust	0.31 (0.03)	64

Notes: The table reports summary statistics on reciprocity behavior by Bs. Column (1) shows the average amount sent back to A as a fraction of the amount received (means with standard errors in parenthesis); Column (2) shows the number of participants B receiving a positive amount in each treatment.

Result 2. *In the No-profit conditions, sending a Trust message does not entail any increase or decrease in the amount of trust reported compared to the Money message.*

Table 2 extends the analysis by comparing the effect of messages across conditions in a Diff-in-Diff approach. The table reports 4 regression models: Models (1) and (2) are OLS regressions analyzing the impact of our treatments on the amount sent, and Models (3) and (4) are Probit models analyzing the impact of the treatments on the decision to send a positive amount. Explanatory variables include two dummy variables for the message treatments (*d(NoMsg)* and *d(MoneyMsg)* with the *Trust* message treatment as the baseline category); a dummy variable for the non-profit treatment *d(NoProfit)*; and their interactions. The combination of our Hypothesis 1a and 1b would predict a reduction in trust between the *Trust* and *Money* message conditions when moving from the



(a) Average amount sent by A (with 95% CI)

(b) Amount sent by A conditional on sending (size of the pie) and fraction of A sending a positive amount (blue share of the pie)

Fig. 1. Profit treatment. Average amount sent by A, fraction of A sending a positive amount to B, and average amount sent by A conditional on sending a positive amount.

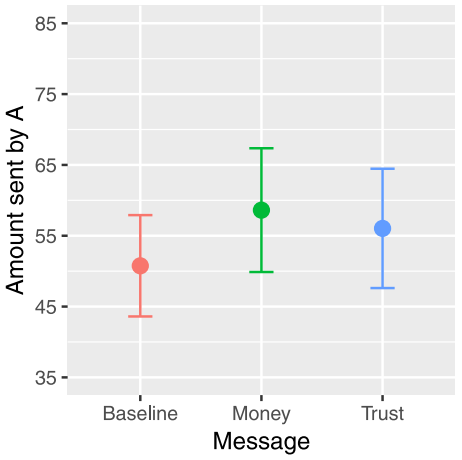
Profit to the *No-profit* setting. In other words, it predicts a negative value for the coefficient associated with the interaction term $d(MoneyMsg) \times d(NoProfit)$.

Regression results show that, although the coefficients are large (equal to about 65% of the effect brought about by sending any message) and in the predicted direction, the standard errors are also large, and the Diff-in-Diff coefficients are consequently not statistically different from zero.

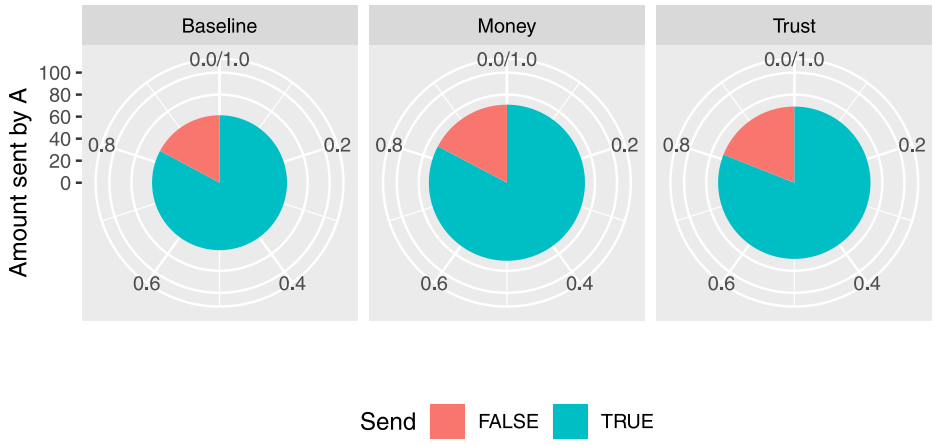
5.3. Analysis of amount reciprocated by the receiver (player B)

We now turn our attention to the receiver's behavior (player B) and discuss the effect of messages on reciprocity. We look at the fraction of the amount received that is sent back to A. Table 3 reports such information split by treatment and message type.

Note that the average fraction sent back is, in general, slightly lower than the break-even amount for player A (i.e., 0.37 in the *For-profit* and 0.33 in the *No-profit* condition). Looking at the effect of messages, we cannot reject the null hypothesis that they have no effect on reciprocity in the *For-profit* condition (One-way anova: $F_{(2,245)} = 1.083$, $p = 0.340$). In the *No-profit* conditions, instead, results are ambiguous (One-way anova: $F_{(2,210)} = 2.999$, $p = 0.052$). Overall, the fractions show a somewhat surprising pattern in the



(a) Average amount sent by A (with 95% CI)



(b) Amount sent by A conditional on sending (size of the pie) and fraction of A sending a positive amount (blue share of the pie)

Fig. 2. No profit treatment. Average amount sent by A, fraction of A sending a positive amount to B, and average amount sent by A conditional on sending a positive amount.

behavior of player B, with the largest fractions returned observed in the *Money* message *No-profit* conditions. Interestingly, there seems to be a mismatch between the behavior of senders and receivers in that it is not the case that more money is sent in the conditions where it is more likely to be returned.

6. Experiment 2: Analysis of attitudes and beliefs

In the main experiment, we found that trust is lower when participant C sends the *Trust* message than when s/he sends the *Money* message in the *For-Profit* condition. At the same time, the *Trust* message has no deflating effects for the amount sent in the *No-profit* condition. With the second experiment, we shed light on the possible explanations for this result by asking a new set of participants questions about their beliefs on the behavior of B in the main task and on the motives driving A's decision to trust.

Several scholarly articles have criticized the use of the rhetoric of sharing by for-profit platforms (Codagnone & Martens, 2016; Eckhardt & Bardhi, 2015; Jemielniak & Przegalinska, 2020). Anecdotal evidence suggests that this criticism may be shared also by the consumer base. For example, Couchsurfing attracted criticism when it became for-profit in 2011 and lost transaction volumes when it started to charge fees to users in 2021 (The Guardian, 2011). At the same time, transactions in the sharing economy are also motivated by strategic considerations and thus shaped by expectations regarding others' behavior. Importantly, the communication

strategy adopted by the platform could alter participants' expectations regarding the intentions and trustworthiness of other platform participants, which could motivate a shift in one's own level of trust.

Experiment 2 aims to measure the role played by the aforementioned psychological and strategic factors in shaping trust in the sharing economy. Specifically, we postulate that the observed treatment effect could be driven by three non-mutually exclusive mechanisms. First, people could share the same opinion reported in the scholarly literature, namely that the rhetoric of sharing to hide traditional corporate objectives of profit maximization is misused and insincere. In this second experiment, we capture this attitude by directly asking participants the degree to which they find the message sent by the platform to be accurate, honest, sincere, and trustworthy.

Second, the communication strategy chosen by the platform could affect participants' higher-order beliefs regarding the intentions and trustworthiness of their fellow platform participants. For example, participants may think that a given communication strategy could lead other participants to act more selfishly or in a less trustworthy manner than another communication strategy, perhaps due to its perceived accuracy or other factors. Lower degrees of trust could then be a rational reaction to lower expectations of trustworthiness. There is ample evidence supporting the relevance of studying expectations of reciprocity in trust games (Evans & Krueger, 2014; Evans, Krueger, et al., 2017; Sofianos, 2022). Since it is player A that faces strategic uncertainty, we capture this channel by measuring participants' beliefs regarding the degree of trustworthiness of player B when receiving different messages. Moreover, we directly ask to report the degree to which a low level of trust by player A could be attributed to the fear of untrustworthiness or a desire on A's part to avoid risk (which may be perceived as higher in the *Trust* message).

The last channel we explore is a desire to punish the platform for adopting a communication strategy that uses the rhetoric of sharing to make profits. We explore this channel by eliciting the degree to which subjects believe that A players "dislike the message sent by C" and "do not want C to profit" from their trust. While this channel is tied to the first, it implies a step forward in that punishment (especially one that comes at a personal cost) does not mechanically follow from finding the message inaccurate or insincere.

6.1. Design and treatments

We ran the second experiment on a new pool of MTurk subjects to avoid contamination across the two experiments. Eliciting beliefs on a subsequent sample of participants, whose only role is answering a belief elicitation task, has become standard in economic experiments (see, for instance, the seminal work by Cohn, Maréchal, Tannenbaum, and Zünd (2019)). After reading the description of the main experiment, participants in this study perform two different tasks in random order.

Task 1. The first task is to guess (a) the number of participants B in the main experiment that sent back a positive amount and (b) the average amount sent back by participants B (among those who sent back a positive amount). We implement this task with a 2 × 2 design. One dimension of variation is between-subjects, and it is the type of message sent by player C (*Trust* vs. *Money* message).[9] The second dimension is within-subjects, which is the amount of money A sends (50 points vs. 100 points). Participants learn that if their guess is accurate, i.e., if the gap between their guess and the actual average behavior of participant B is smaller than 5, they obtain a bonus payment. Thus, overall, each participant makes four guesses: the number of participants B sending back a positive amount, the average amount sent back conditional on A sending 50 points, and the same two measures conditional on A sending 100 points.

Task 2. The second task elicits the shared opinion regarding (a) the message sent by C and (b) the reasons why As do not send all the points to Bs. For this purpose, we use the following questions:

1. *What do you think about the message sent by C?* [From 1 (Strongly disagree) to 7 (Strongly agree)].
    - The message sent by C is trustworthy
    - The message sent by C is honest
    - The message sent by C is sincere
    - The message sent by C is accurate
2. *Why do you think participant As did not send all the points they had to B?* [From 1 (not at all important) to 7 (very important)]
    - because As fear B would keep all the money
    - because As do not want to take risks
    - because As do not like the message sent by C
    - because As do not want C to profit from them sending points

We incentivize these questions with a modification to the Krupka–Weber method (Krupka & Weber, 2013): subjects have to guess the mean of the answers given by all the participants in the study. For each of the eight statements, we assigned a bonus payment to the participant whose guess is the closest to the actual mean. This way, we incentivize the accurate guess of the shared opinion.

Instead of incentivizing the guessing of the median, as in the original study by Krupka and Weber (2013), we incentivized the guessing of the mean. We are not the first to do so (see, for instance, Schmidt, Heinicke, and König-Kersting (2022)). Prior studies

[9] As for Experiment 1, in Experiment 2 participants were not aware of the possibilities available to player C. This makes sure that participants in this experiment can form beliefs that are comparable to those of participants A and B in Experiment 1.

show that the concept of the median is more challenging to understand than the mean (Zawojewski & Shaughnessy, 2000). The understanding of the task is essential in our online setting of MTurk workers, whose attention levels tend to be lower than in standard lab experiments and in which there is no opportunity for in-person personalized explanations. The applied technique is still incentive-compatible. We decided to trade lower robustness to outliers with a higher comprehension of the instructions.

The between-subject treatment manipulation is the type of message sent by C, i.e., either the *Trust* or the *Money* message. The message faced by the subjects in task one is the same as faced in task two. We provide a copy of the instructions in the Online Appendix.

6.2. Pre-registered hypotheses

Based on the motivation highlighted at the beginning of this section, we outline here our formal and pre-registered hypotheses.[10]

Our first hypothesis relates to the first channel discussed above, namely, that the reduction of trust caused by the *Trust* message is driven by a perception of dishonesty, deceitfulness, or inaccuracy of this message compared to the *Money* message. We therefore formalize the following hypothesis:

Hypothesis 2a (*Perceived Honesty of the Messages*)**.** If the effect of *Trust* vs. *Money* message obtained in Study 1 is driven by a perception of dishonesty, inaccuracy, or insincerity, we expect the *Trust* message to obtain lower scores than the *Money* message in the answers to Question 1 of Task 2.

Our next hypothesis is based on the second channel identified in the discussion above, namely, that the effect of the message is to change the beliefs about the reciprocity of player B. Here we expect the following:

Hypothesis 2b (*Reduced Expectation of Reciprocity*)**.** If the effect of *Trust* vs. *Money* message obtained in Study 1 is driven by A's beliefs that the *Trust* message reduces B's reciprocation, we expect:

(i) in Task 1, more pessimistic beliefs about the number of Bs reciprocating and about the level of reciprocity with the *Trust* message compared to the *Money* message;
(ii) in Question 2 of Task 2, higher scores to the answers "because As fear B would keep all the money" and "because As do not want to take risks" with the *Trust* compared to the *Money* message.

Our last hypothesis is based on the third channel identified in the discussion above, namely, that the effect of the message is to trigger an aversion towards player C's attempt to use a trust rhetoric to increase profits. We hypothesize that:

Hypothesis 2c (*Aversion Towards C*)**.** If the effect of *Trust* vs. *Money* message obtained in Study 1 is driven by an aversion towards player C's use of the trust rhetoric, we expect:

(i) in Task 1, no difference in the beliefs about the number of Bs reciprocating and about the level of reciprocity with the two messages;
(ii) in Question 2 of Task 2, higher scores to the answers "because As do not like the message sent by C" and "because As do not want C to profit from them sending points" with the *Trust* compared to the *Money* message.

6.3. Participants and procedures

We ran the second experiment in December 2020 with a total of 403 subjects completing the questionnaire (202 in *Trust* and 201 in *Money*). To ensure comparability with the main experiment, the data was collected on MTurk using the same restrictions we used in the previous study. Notably, we excluded workers who took part in the main study. Participants earned a participation fee of $0.50 upon completion of the study; a bonus of $0.15 per 1 correct guess in the first task, and a bonus of $1.00 if they provided the most accurate guess for one of the statements in the second part.

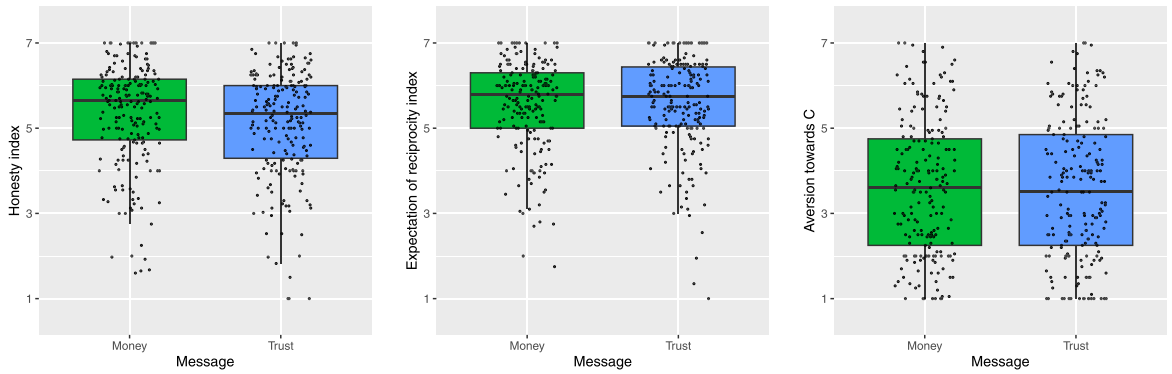
6.4. Results

As specified in our pre-registration, to test Hypothesis 2a, we compute the average score of the answer to Question 1. We label this index "perceived honesty" of the message. Overall, Panel (a) of Fig. 3 shows that the perceived honesty indexes of the *Money* message are higher than the ones of the *Trust* message (Mann–Whitney test: $p = 0.042$). This supports the idea that participants in the *For-profit* condition may perceive as less honest a rhetoric of trust compared to a rhetoric of profit.[11]

Hypothesis 2b requires testing two sub-hypotheses. First, we test whether the participant's beliefs regarding reciprocity of B players are more pessimistic with the *Trust* message compared to the *Money* message. Then, we need to test whether the motives capturing a low expectation of reciprocity – i.e., the ratings of the statements "because As fear B would keep all the money" and

[10] Pre-registration link: https://osf.io/8udhp/files/osfstorage/5fcf9d456ebcc6040746a921

[11] Testing separately each of the dimensions of Question 1 provides the following results: Trustworthy, $p = 0.087$; Honest, $p = 0.077$; Sincere, $p = 0.214$; Accurate, $p = 0.007$.



(a) Perceived honesty of the message (b) Expectation of reciprocity (c) Aversion towards C

Fig. 3. Answers to Question 1 and Question 2.

"because As do not want to take risks" in Question 2 – are stronger in the *Trust* compared to the *Money* message. Fig. 4 shows the beliefs of how many participants B reciprocate out of 100 (in panel (a)) and of how much participants B return to A (in panel (b)). We cannot reject the null hypothesis of no effect of the message for all the comparisons.[12] As pre-registered, for the expectation of reciprocity, we computed an index averaging the ratings of the two statements "because As fear B would keep all the money" and "because As do not want to take risks". The distribution of the indexes of expectation of reciprocity is reported in Panel (b) of Fig. 3.[13] The Figure shows how participants have similar ratings of the expectation of reciprocity after reading the two messages (Mann–Whitney test: $p = 0.850$).

Overall, the two results provide no support for Hypothesis 2b. Subjects in our experiment do not believe that player A's motivation not to trust stems from an expectation that Bs reciprocate less after viewing the *Trust* message compared to the *Money* message.

As for our Hypothesis 2c, this, too, requires testing two sub-hypotheses. The first is that there are no differences in the expectation of reciprocity, which we already established testing Hypothesis 2a; the second is that participants believe that A's decisions are motivated by a higher aversion towards C in the *Trust* message compared to the *Money* message. To test the latter, we follow the pre-registration and compute an index of "Aversion towards C" by averaging the ratings of the two statements: "because As do not like the message sent by C" and "because As do not want C to profit from them sending points". The distributions of the index of aversion towards C for the two messages, reported in Panel (c) of Fig. 3, show that participants do not believe that A has perceived a higher aversion towards C when exposed to the *Trust* message (Mann–Whitney test: $p = 0.876$).

After testing our hypotheses, we can summarize the outcome of the second experiment in the last result of our study:

Result 3. *In the For-profit conditions, the driver for more subjects sending a positive amount after the Money message compared to the Trust message is the perception of the Money message being more honest.*

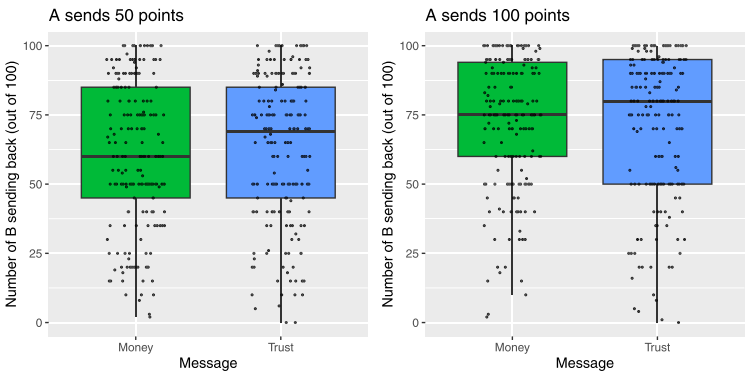
Overall, Experiment 2 advances our understanding of the factors behind the result of Experiment 1, primarily by rejecting our hypothesized mechanisms. Aside from finding indications for an effect of the "perceived honesty" of the message, the experiment highlights factors that do not seem to explain our main results.

One possible explanation for the findings is that the main treatment effect is due to other factors that we did not consider. The fact that the *Trust* message is perceived as a more dishonest and inaccurate description but does not influence the expected reciprocity or the attitudes towards the platform is suggestive of a more subtle channel through which rhetoric can impact trust. One possible speculation is that the contrast between the message and the reward structure of the platform could generate some cognitive dissonance which, in turn, may reduce the willingness to trust (without impacting the measured attitudes and higher-order beliefs). The link between trust, cognitive dissonance, and brand affiliation has been documented, for example, by Shahin Sharifi and Rahim Esfidani (2014).

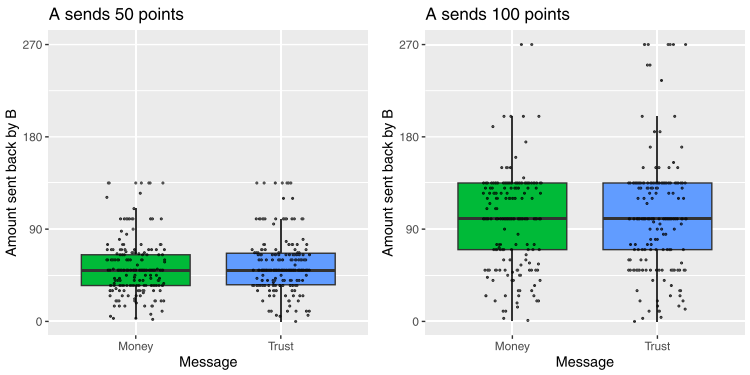
Another possibility is that the experimental design failed to capture the role played by these factors. This could be due to the complexity of the experiment whereby observers are asked to form hypothetical expectations about how other players have played the game (Hauge, Brekke, Johansson, Johansson-Stenman, & Svedsäter, 2016). This could have generated some noise in the responses, which can introduce some power issues. Future research is needed to disentangle these different hypotheses and further explore the underpinning of the behavioral shift observed in Experiment 1.

[12] Mann–Whitney tests: Number of B sending back when A sends 50: $p = 0.342$; Number of B sending back when A sends 100: $p = 0.943$; Amount sent back when A sends 50: $p = 0.117$; Amount sent back when A sends 100: $p = 0.852$.

[13] Note that this index is reverse coded, with higher levels of the index implying lower expectation of reciprocity.



(a) Beliefs about the number of B reciprocating (sending back a positive amount)



(b) Beliefs about the amount sent back by the Bs that reciprocate

Fig. 4. Beliefs about reciprocity of B.

7. Conclusions

Trust is essential in virtually all economic and social dynamics, especially in the context of the sharing economy (Ben-Ner & Halldorsson, 2010). In this work, we were interested in the following questions: Does the revenue model of a sharing platform interact with the communication strategy of the platform in shaping trust? More specifically, is a rhetoric of sharing successful when the promoters benefit financially from it? Our findings support a negative answer to the latter question.

We modify a standard trust game in an online experiment by introducing a third player. Its role is to decide whether or not to send a pre-formulated message, which, depending on the treatments, can be either a *Money* message that stresses the monetary rewards accruing from trusting or a *Trust* message that evokes positive feelings of trust and reciprocity.

Furthermore, third players' monetary rewards change across treatments: in the *No-profit* condition, aimed to mimic pure sharing platforms, the third player earns a flat fee that does not depend on the amount of trust generated in the interaction between the trustor and the trustee. Conversely, in the *For-profit* condition, which aims to resemble lucrative sharing platforms, the third player earns a fee proportional to the amount of trust generated in the trust game.

Our main result is that, in the *For-profit* conditions, there is a lower frequency of trustors sending a positive amount to the trustee after receiving a *Trust* message than a *Money* message. When, in a second experiment, we investigate the reasons underlying this finding, we observe that the driver is the belief of the *Money* message being more accurate and honest compared to the *Trust* message. The second experiment provides an interesting viewpoint on the drivers of this adverse effect. This is done by comparing the trustee's behavior and the beliefs about the amount sent back. Indeed, while trustors send different amounts when exposed to the two messages, the amount returned by trustees and the participants' beliefs about this amount are not influenced by the message. The two results together suggest that the lower trust observed when the third player uses a rhetoric of trust is not driven by A's beliefs about B's reaction to the message but rather by the low accuracy and honesty of the message sent, originating from the discrepancy between the nature of the message and the nature of the interaction.

The main behavioral mechanism thus seems to hinge on the connection between perceived honesty and trust generation. In stylized settings, Rotenberg, MacDonald-Taylor, and Holland (2023) have found that consistency between personal standards and behavior is a strong predictor of how trustworthy an agent is perceived to be, while Bellucci and Park (2020) show that a perception

of honesty produces trust so strongly that it persists even when the other party has behaved untrustworthily. In the marketing literature, brand authenticity (i.e., the degree to which the communication of a company aligns with its products and actions) has repeatedly been established as a key driver of trust and loyalty towards a brand Hernandez-Fernandez and Lewis (2019), Portal, Abratt, and Bendixen (2019), Södergren (2021).

Our study contributes to understanding possible mechanisms to boost trust in online sharing environments, which rely on trust generation as their core feature. This is especially relevant in a time in which the *marketization* process has been documented to be a predictor of trust decline (Xin & Xin, 2017). Our evidence suggests that utilizing the rhetoric of sharing by profiting platforms can have detrimental effects on trust generation. Our results provide some scientific background for the backlash faced by sharing economy platforms that started as pure-sharing platforms and tapped heavily on the rhetoric of trust, only to switch to a lucrative business model some years later (Schor & Cansoy, 2019). Our findings imply that providing sharing platforms should go hand-in-hand with accurate and sincere communication.

One limitation of our study is that, although we find evidence for our main result in the form of a simple effect, this is not corroborated in terms of statistical significance by a Diff-in-Diff approach performed interacting the two dimensions of treatments (*Profit* or *No-profit* and *Money* or *Trust* message). Moreover, the analysis conducted in the second experiment, investigating the roots of possible mechanisms underlying our results, lends support only to the perception of the *Money* message sent in *For-profit* conditions to be more accurate and honest with respect to the *Trust* message.

With the growing popularity of sharing platforms, we see this study as a first step to understanding the interaction between communication strategies and earning models in shaping trust. However, many open questions remain, including the investigation of hybrid communication strategies and the dynamic effects of changes to both the earning models and the communication strategies within ongoing partnerships. One interesting avenue of future research could examine how the effect of rhetoric changes when the reward structure is chosen and not exogenously set by the experimenters, as in our setting. Our conjecture is that subjects would have punished the platform more for deliberately selecting a for-profit reward structure.

Data availability

The data is publically available via OSF (https://osf.io/8udhp/?view_only=).

Acknowledgments

We would like to thank Margarita Leib, Nils Kobis, Shaul Shalvi, Jeroen Van Der Ven, Joel Van Der Weele, and seminar participants at CREED, Leiden University, M-BEES Maastricht workshop, ESA Global conference and CIMEO workshop on Experimental Economics (Rome, University La Sapienza). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 865931).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.joep.2024.102728.

References

Banuri's, S., de Oliveira, A. C., & Eckel, C. C. (2019). Care provision: An experimental investigation. *Journal of Economic Behaviour and Organization*, *157*, 615–630.
Belk, R. (2014a). Sharing versus pseudo-sharing in web 2.0. *The Anthropologist*, *18*(1), 7–23.
Belk, R. (2014b). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, *67*(8), 1595–1600.
Bellucci, G., & Park, S. Q. (2020). Honesty biases trustworthiness impressions. *Journal of Experimental Psychology: General*, *149*(8), 1567–1586.
Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79.
Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.
Bicchieri, C., Duffy, J., & Tolle, G. (2004). Trust among strangers. *Philosophy of Science*, *71*(3), 286–319.
Bigoni, M., Bortolotti, S., Casari, M., & Gambetta, D. (2013). It takes two to cheat: An experiment on derived trust. *European Economic Review*, *64*, 129–146.
Botsman, R. (2013). The sharing economy lacks a shared definition. *Fast Company*, *21*, 2013.
Brandts, J., Cooper, D. J., & Rott, C. (2019). Communication in laboratory experiments. *Handbook of Research Methods and Applications in Experimental Economics*, *401*.
Cadsby, C. B., Maynes, E., & Trivedi, V. U. (2006). Tax compliance and obedience to authority at home and in the lab: A new experimental approach. *Experimental Economics*, *9*, 343–359.
Capraro, V., Vanzo, A., & Cabrales, A. (2021). Playing with words: Do people exploit loaded language to affect others' decisions for their own benefit? *Judgment & Decision Making*.
Cassar, A., & Rigdon, M. (2011). Trust and trustworthiness in networked exchange. *Games and Economic Behavior*, *71*(2), 282–303.
Chang, D., Chen, R., & Krupka, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, *116*, 158–178.
Charness, G., & Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, *107*(2), 281–283.
Chaudhuri, A., Li, Y., & Paichayontvijit, T. (2016). What's in a frame? Goal framing, trust and reciprocity. *Journal of Economic Psychology*, *57*, 117–135.
Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.
Chen, Y., & Wang, L. (2019). Commentary: Marketing and the sharing economy: Digital economy and emerging market challenges. *Journal of Marketing*, *83*(5), 28–31.
Cheng, X., Fu, S., Sun, J., Bilgihan, A., & Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management*, *71*, 366–377.

Chica, M., Chiong, R., Adam, M. T., & Teubner, T. (2019). An evolutionary game model with punishment and protection to promote trust in the sharing economy. *Scientific Reports*, *9*(1), 1–12.
Codagnone, C., & Martens, B. (2016). Scoping the sharing economy: Origins, definitions, impact and regulatory issues. Available at: https://www.econstor.eu/bitstream/10419/202218/1/jrc-dewp201601.pdf (Accessed August 2021).
Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, *365*(6448), 70–73.
Coleman, J. S. (1990). *Foundations of social theory*. Harvard University Press.
Corten, R. (2019). Social dilemmas in the sharing economy. In *Handbook of the sharing economy*. Edward Elgar Publishing.
Cronk, L. (2007). The influence of cultural framing on play in the trust game: A Maasai example. *Evolution and Human Behaviour*, *28*(5), 352–358.
Deutsch, M. (1958). Trust and suspicion. *The Journal of Conflict Resolution*, *2*(4), 265–279.
Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, *13*(2), 123–139.
Dunning, D., Fetchenhauer, D., & Schlösser, T. M. (2012). Trust as a social and emotional act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology*, *33*(3), 686–694.
Eckhardt, G. M., & Bardhi, F. (2015). The sharing economy isn't about sharing at all. *Harvard Business Review*, *28*(1), 881–898.
Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, *76*(1), 117–130.
Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2013). Gender differences in social framing effects. *Economics Letters*, *118*(3), 470–472.
Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment & Decision Making*, *9*(2).
Evans, A. M., Krueger, J. I., et al. (2017). Ambiguity and expectation-neglect in dilemmas of interpersonal trust. *Judgment & Decision Making*, *12*(6), 584–595.
Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, *7*(2–3), 235–266.
Fréchette, G. R., Sarnoff, K., & Yariv, L. (2022). Experimental economics: Past and future. *Annual Review of Economics*, *14*, 777–794.
Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The sharing economy: Why people participate in collaborative consumption. *Journal of the Association for Information Science and Technology*, *67*(9), 2047–2059.
Hauge, K. E., Brekke, K. A., Johansson, L.-O., Johansson-Stenman, O., & Svedsäter, H. (2016). Keeping others in our mind or in our heart? Distribution games under cognitive load. *Experimental Economics*, *19*, 562–576.
Hauser, D., Paolacci, G., & Chandler, J. (2019). Common concerns with MTurk as a participant pool: Evidence and solutions. In *Handbook of research methods in consumer psychology* (pp. 319–337). Routledge.
Hawlitschek, F., Teubner, T., & Weinhardt, C. (2016). Trust in the sharing economy. *Die Unternehmung*, *70*(1), 26–44.
Hernandez-Fernandez, A., & Lewis, M. C. (2019). Brand authenticity leads to perceived value and brand trust. *European Journal of Management and Business Economics*, *28*(3), 222–238.
Hofmann, E., Hoelzl, E., Sabitzer, T., Hartl, B., Marth, S., & Penz, E. (2022). Coercive and legitimate power in the sharing economy: Examining consumers' cooperative behavior and trust. *Journal of Economic Psychology*, *93*, Article 102565.
Hong, K., & Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, *28*(2), 197–213.
Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*, 399–425.
Jemielniak, D., & Przegalinska, A. (2020). *Collaborative society*. MIT Press.
Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889.
Karakostas, A., & Zizzo, D. J. (2016). Compliance and the power of authority. *Journal of Economic Behaviour and Organization*, *124*, 67–80.
Ko, G., Amankwah-Amoah, J., Appiah, G., & Larimo, J. (2022). Non-market strategies and building digital trust in sharing economy platforms. *Journal of International Management*, *28*(1), Article 100909.
Koopman, C., Mitchell, M., & Thierer, A. (2014). The sharing economy and consumer protection regulation: The case for policy change. *Journal of Business, Entrepreneurship & the Law*, *8*, 529.
Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*(3), 495–524.
Leonard, L. N. (2012). Attitude influencers in C2C e-commerce: Buying and selling. *Journal of Computer Information Systems*, *52*(3), 11–17.
Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, *30*(9), 1175–1185.
Lundquist, T., Ellingsen, T., Gribbe, E., & Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behaviour and Organization*, *70*(1–2), 81–92.
McKnight, D. H., & Chervany, N. L. (2001). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, *6*(2), 35–59.
Pelligra, V., Reggiani, T., & Zizzo, D. J. (2020). Responding to (un) reasonable requests by an authority. *Theory and Decision*, *89*(3), 287–311.
Phua, V. C. (2019). Perceiving Airbnb as sharing economy: the issue of trust in using Airbnb. *Current Issues in Tourism*, *22*(17), 2051–2055.
Portal, S., Abratt, R., & Bendixen, M. (2019). The role of brand authenticity in developing brand trust. *Journal of Strategic Marketing*, *27*(8), 714–729.
PWC (2015). The sharing economy. https://www.pwc.com/us/en/technology/publications/assets/pwc-consumer-intelligence-series-the-sharing-economy.pdf. (Last Accessed 9 February 2023).
Rietz, T. A., Sheremeta, R. M., Shields, T. W., & Smith, V. L. (2013). Transparency, efficiency and the distribution of economic welfare in pass-through investment trust games. *Journal of Economic Behaviour and Organization*, *94*, 257–267.
Rotenberg, K. J., MacDonald-Taylor, B., & Holland, R. (2023). Trusting others who vary in consistency between their personal standards and behavior: Differences by age, gender, and honesty trust beliefs.. *Social Development*.
Schmidt, R., Heinicke, F., & König-Kersting, C. (2022). Using coordination games to measure beliefs. *Economics Letters*, *219*, Article 110821.
Schor, J. B., & Cansoy, M. (2019). The sharing economy. *The Oxford Handbook of Consumption*, 51–74.
Schor, J. B., & Vallas, S. P. (2021). The sharing economy: Rhetoric and reality. *Annual Review of Sociology*, *47*.
Shahin Sharifi, S., & Rahim Esfidani, M. (2014). The impacts of relationship marketing on cognitive dissonance, satisfaction, and loyalty: The mediating role of trust and cognitive dissonance. *International Journal of Retail and Distribution Management*, *42*(6), 553–575.
Sheremeta, R. M., & Zhang, J. (2014). Three-player trust game with insider communication. *Economic Inquiry*, *52*(2), 576–591.
Södergren, J. (2021). Brand authenticity: 25 years of research. *International Journal of Consumer Studies*, *45*(4), 645–663.
Sofianos, A. (2022). Self-reported & revealed trust: Experimental evidence. *Journal of Economic Psychology*, *88*, Article 102451.
Sonntag, A., & Zizzo, D. J. (2015). Institutional authority and collusion. *Southern Economic Journal*, *82*(1), 13–37.
Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, *8*, 321–340.
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.
The Guardian (2011). Not-for-profit couchsurfing becomes a company (with a conscience). https://www.theguardian.com/travel/2011/aug/26/couchsurfing-investment-budget-travel. (Last Accessed 9 February 2023).
Vaughan, R., & Hawksworth, J. (2014). The sharing economy: How will it disrupt your business. *Megatrends: The collisions. PwC Presentation*.
Wuthisatian, R., Pingle, M., Nichols, M., et al. (2017). To support trust and trustworthiness: punish, communicate, both, neither? *Journal of Behavioral Economics for Policy*, *1*(1), 61–68.
Xin, Z., & Xin, S. (2017). Marketization process predicts trust decline in China. *Journal of Economic Psychology*, *62*, 120–129.
Zawojewski, J. S., & Shaughnessy, J. M. (2000). Take time for action: Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School*, *5*(7), 436–440.