Games and Economic Behavior 139 (2023) 15–25

Contents lists available at ScienceDirect

Games and Economic Behavior

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

Honesty in the city

Martin Dufwenberg [a,b,c], Paul Feldman [d], Maroš Servátka [e,f,*], Jorge Tarrasó [g], Radovan Vadovič [h]

[a] University of Arizona, United States of America
[b] University of Gothenburg, Sweden
[c] CESifo, Germany
[d] Texas A&M University, United States of America
[e] MGSM Experimental Economics Laboratory, Macquarie Business School, Australia
[f] University of Economics in Bratislava, Slovakia
[g] Libretto, United States of America
[h] Carleton University, Canada

ARTICLE INFO

Article history:
Received 15 May 2021
Available online 31 January 2023

JEL classification:
C72
C90
C93
D91

Keywords:
Trustworthiness
Honesty
Reciprocity
Field experiment
Haggling
Taxis
Mexico City

ABSTRACT

Lab evidence on trust games involves more cooperation than conventional economic theory predicts. We explore whether this pattern extends to a field setting where we are able to control for (lack of) repeat-play and reputation: the taxi market in Mexico City. We find a remarkably high degree of trustworthiness, even with price-haggling which was predicted to reduce trustworthiness.

1. Introduction

During the last quarter-century, economists argued that social preferences shape behavior in important ways. Many laboratory studies conducted with students (who possess qualities researchers deem convenient like being accessible and motivated by low stakes) artificially create anonymous settings that rule out repeated play and reputation building as potential confounds. Over time, the accumulated lab evidence confirms that trust, cooperation, and honesty are abundant but not universal.

* Corresponding author.
E-mail addresses: martind@eller.arizona.edu (M. Dufwenberg), paul.feldman@ag.tamu.edu (P. Feldman), maros.servatka@mgsm.edu.au (M. Servátka), jorge.tarraso@gmail.com (J. Tarrasó), radovan.vadovic@carleton.ca (R. Vadovič).

https://doi.org/10.1016/j.geb.2023.01.007
0899-8256/

It is natural to ponder whether a similar pattern occurs in the field,[1] and stylized trust games have been used also outside the lab.[2] But when it comes to naturally occurring trust situations, it is a challenge to maintain control for repeat-play and reputation. We overcome that hurdle by conducting an experiment in the large and highly decentralized market for taxi rides in Mexico City, where the chances of repeated encounters are minuscule.[3] Beyond learning about a specific setting and location, our study contributes a method to explore other field settings.

Our primary focus is trustworthiness. We flag down cabs at point A, ask them to deliver an item that has value for the customer at point B, and pay in advance. From the viewpoint of the driver, the situation is comparable to that of the second-mover (the "trustee") in a trust game, in particular versions that allow for communication before play.[4] In the lab, trustees cooperate with high but less than full frequency. We explore whether that pattern has an analog in the streets of Mexico City.

We ran a pilot for that Baseline treatment, planning to condition further research questions and treatments on the nature of the data. If trustworthiness were low, we would have a treatment adding handshakes & promises to the pre-play communication with the driver to test if such enhanced covenants boost trustworthiness. However, trustworthiness was extremely high already in the Baseline, so we scrapped the promises & handshakes treatment. Instead, we ran a Haggling treatment, predicted to instead reduce trustworthiness: If a cab driver asked for a price p then we haggled and tried to bring the price down to $p_r < p$. Our intuitive hypothesis, which (as we show) is also consistent with reciprocity theory, is that delivery rates will be lower in Haggling than in Baseline.

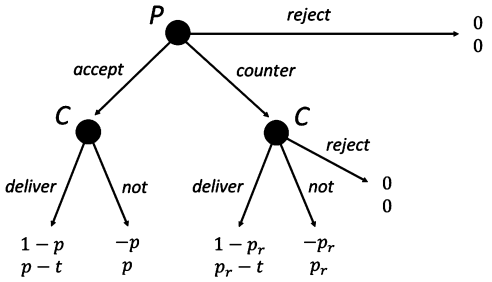
We also explore two robustness conditions, involving (i) a nearby but less prominent route and (ii) the original route but with data collected several years later.[5] We postpone a more detailed discussion to Section 3.

Some former bargaining studies document whether parties are willing to deceptively use unverifiable private information to secure a larger share of resources. Analogous questions are asked in studies exploring credence goods markets where an uninformed party may get overcharged or subjected to more treatment than warranted.[6] Our focus is not on informational asymmetries, but rather on trustworthiness and how haggling affects the delivery rate.[7] Our work also relates to the larger literature on honesty and cheating, often found to be sensitive to a variety of incentives and contexts.[8]

Section 2 describes our setting theoretically, and the hypothesis we test. Sections 3 and 4 describe our experimental design and results. Section 5 concludes. An appendix contains the proof of a proposition, figures with data, maps with the routes, and results from a survey with locals.

2. Theory

Assume that cab driver C perceives that he interacts with passenger P as follows:



[1] Studies with a related motivation to test whether phenomena observed in the lab also occur (with a similar intensity) in the field include exploration of peer punishment (Balafoutas and Nikiforakis, 2012), and whether people return a lost wallet (Cohn et al., 2019).

[2] For instance, Falk and Zehnder (2013) study the effect of in-group favoritism on trust in a city-wide experiment in Zurich while Finseraas et al. (2019) examine the impact of ethnic diversity among Norwegian soldiers.

[3] According to Mexico's "Secretaria de Transportes y Vialidad" in 2011, 102,110 licensed taxis served almost 9 million people in Mexico City, and over 21 million in the larger metropolitan area. The taxis drivers are self-proprietors holding individual licenses (as opposed working for a taxi company) making it virtually impossible to track them down.

[4] See, e.g., Berg et al. (1995) for a pioneering contribution and Charness and Dufwenberg (2006) and Vanberg (2008) for versions with pre-play communication.

[5] Our original experiment was carried out in 2011, before the ride-sharing cell-phone technology enabled passengers to track the drivers. The additional session run in 2022 took place after the entry of Uber and other ride-sharing providers to the Mexico City taxi market.

[6] See, e.g., laboratory experiments by Huck et al. (2007), Dulleck et al. (2011), Huck et al. (2012), and field experiments by Schneider (2012), Balafoutas et al. (2013), and Gottschalk et al. (2020); for a survey, see Kerschbamer and Sutter (2017).

[7] The impact of negotiations on the economic outcomes is also considered by Bhattacharya and Dugar (2020) who find that sellers in a fish market are more likely to cheat on the weight following a price negotiation instigated by the buyer. They differ from us in allowing for repeat-play, not having any opportunity for the seller to reneg, and in that their customers may not find out whether they were cheated.

[8] See, e.g., Gneezy (2005), Dreber and Johannesson (2008) Ellingsen et al. (2009), Lundquist et al. (2009), Sutter (2009), Houser et al. (2012), Azar et al. (2013), Fischbacher and Föllmi-Heusi (2013), Gibson et al. (2013), Innes and Mitra (2013), Pruckner and Sausgruber (2013), Abeler et al. (2019), Dugar and Bhattacharya (2017), Gneezy et al. (2018), Dugar et al. (2019). A few other recent studies of taxi markets have also addressed questions of cooperation and cheating, e.g., Balafoutas et al. (2013, 2017); Castillo et al. (2013); Bengtsson (2016).

C has offered to deliver a valuable item for the price p. In response, P may *reject* (and walk away) or *accept* or *counter* at the price $p_r < p$. In the latter two cases, C may choose whether to *deliver* or *not*, and in response to *counter*, C may furthermore also *reject* (and walk away). We normalize payoffs such that each player gets 0 if either player *rejects* and P's value of a safe delivery equals 1. In addition, C faces transportation costs of t. To allow meaningful gains-from-trade, assume that $1 > p > p_r > t > 0$.

We now have a form of trust game. If the players are selfish then there is a unique subgame perfect equilibrium: C chooses not to deliver at each of his nodes and P rejects at the root. In analogy with observations from lab trust games one may, however, suspect that C and P will behave differently. In our experiment, we explore whether this is the case. We furthermore find it intuitive that if one player haggles then the other may become less trustworthy. Accordingly, we test the hypothesis that C's delivery rate is lower if P chooses *counter* rather than *accept*.

One way to formally justify that prediction is to assume that players are motivated by kindness-based reciprocity, as we now show using Dufwenberg and Kirchsteiger's (2004) (D&K) notion of sequential reciprocity equilibrium (SRE)[9]: C's choices as a behavior strategy: Let δ be the probability of *deliver* following *accept*. Let ε be the probability of *deliver* following *counter*. Let ρ be the probability of *reject* following *counter*. Parameter $Y_{CP} \geq 0$ measures the degree to which C derives utility from reciprocation, with higher values meaning higher sensitivity. We have:

Proposition 1. (i) *In any SRE it holds that* $\delta \geq \varepsilon \geq \rho = 0$. (ii) *There exist* $h > \ell > 0$ *such that if* $\ell < Y_{CP} < h$ *then in any SRE it holds that* $\delta > \varepsilon$.

The proof requires an introduction of D&K's formalism, which we provide in the Appendix. Here we just offer the following brief interpretation: In SRE, C always perceives *accept* by P to be at least as kind as *counter*, so C's inclination to be kind in return is reflected in that $\delta \geq \varepsilon$. Moreover, if C is eager to reciprocate kindness but not too eager ($\ell < Y_{CP} < h$) then C is strictly more likely to deliver following *accept* than *counter* ($\delta > \varepsilon$).

3. Experimental design

For our main treatments, we selected a six-mile route along Vía Insurgentes, a straight and major traffic artery in Mexico City. A cab ride takes 20-30 minutes and costs about 30 pesos. We used so-called "de calle" taxis that free-lance around the city. The cab drivers are self-proprietors and are not organized nor use any radio/phone/internet operated service. We employed two research assistants (RA) who were native speakers. The RA at point A ("RA-A") flagged down a cab, and asked the driver to deliver a CD containing a movie clip to his friend ("RA-B") at point B. RA-A asked for the price, explained that the friend at point B had no money, and proposed to pay up front. The next move depended on the treatment.

The cab drivers were assigned to treatments – Baseline or Haggling. In Baseline, RA-A agreed to the offered fare. In Haggling, RA-A made a counteroffer, subtracting 10 pesos from the driver's proposal. If the driver made a counteroffer, RA-A agreed and paid up. If the driver rejected the lower fare, then RA-A tried to get a discount by making successive offers until the driver agreed.[10] (There were no rejections.) After a deal was struck, RA-A thanked the driver, described what the friend looked like, and paid. As the cab drove off, RA-A discretely recorded the data and sent an SMS to RA-B. Once the cab arrived at point B, RA-B collected the item (CD in 2011, USB in 2022) and thanked the driver. If the driver demanded additional money, RA-B pretended he did not know that the fare has already been paid, paid the driver, and recorded the sum, the license plate, and the time (to identify whether the delivery was made and by which cab).

4. Results

Seven cabs refused to be hired. Among the others, we count 31 observations in Baseline and 30 in Haggling.[11] For those, the initial price proposals averaged 38.97 in Baseline[12] and 40.00 in Haggling and were not significantly different ($p = 0.546$; Mann-Whitney ranksum test). The final prices in Haggling averaged 32.70 and were significantly lower than initial proposals in either treatment ($p < 0.001$ for both comparisons). There is no statistical difference in the amount of time it took the cabs to deliver the CD between the two treatments ($p = 0.359$).

Our two main findings are as follows:

[9] D&K extend to extensive games ideas about reciprocity pioneered by Rabin (1993). For broader discussions about reciprocity, see, Fehr and Gächter (2000), Sobel (2005), and Battigalli and Dufwenberg (2022, Section 2).

[10] Neither sending packages using cabs nor haggling ex ante over the fare is the norm in Mexico City. One typically gets into the cab and pays the meter rate.

[11] RA-A felt uncomfortable haggling with drivers who made very low initial offer ($\leq$25 Pesos) and assigned such cabs to Baseline. He similarly assigned an unusually high offer ($\geq$60 Pesos) to Haggling. To eliminate possible selection bias at the tails of our initial offer distributions we excluded those data points and conduct the analysis on the remaining 61 observations. Including the left-out observations does not change our main results.

[12] In Baseline, two cab drivers refused to make a proposal, instead asking RA-A for an offer. RA-A guessed an average fare based on traffic conditions. In one case he offered 30, in the other 50. Both offers were accepted (and the observation assigned to Baseline).

Table 1
Other deviations from the agreement.

	Observations	Asked for extra $	Asked to cover the cost
Baseline	31	3	3
Haggling	30	6	5
AltRoute	20	3	2
Post-Uber	32	4	0

- Comparing our setting to all those lab experiments with students (cited in the introduction), our measure of trustworthiness (delivery rates) is *much* higher. With the exception of a single cab driver (in BASELINE), every cab driver who agreed to deliver the CD did so.
- The prediction that delivery rates would be lower following price haggling was not supported; see the previous bullet.[13]

We also offer the following complementary observations:

What did locals expect?

We didn't in our wildest dreams predict the astonishing level of honesty (as regards delivery rates) exhibited by the cab drivers. On seeing the data, we got curious whether also locals would find the results surprising. Therefore, we conducted a survey with students at Instituto Tecnológico Autónomo de México (ITAM). We used the technique for eliciting shared opinions developed by Houser and Xiao (2011). Many locals seem likeminded to us. We did not find any evidence of consensus between them that all cabs would deliver. See the Appendix for details.

Robustness check #1: alternative route

We replicated the procedures of the HAGGLING treatment except that we operated on a less prominent but nearby route (Calle Bolivar, between Avenida Hidalgo and Eje 6 Sur; see the Appendix for a map). One may wonder whether taxi drivers are more likely to renege when operating on street which is less in the spotlight. We ran the HAGGLING treatment with 20 cabs. There was no difference between this sample and the original sample collected for the HAGGLING treatment in terms of offers or trip durations.[14] Just like in the HAGGLING treatment, *all* cabs delivered the CD to point B.

Robustness check #2: Post-Uber market

Our original treatments were run before the market entry of ride-sharing providers, like Uber or Didi. These new services allow for car-tracking, which may discourage drivers from engaging in confidence trickstery. One may conjecture that untrustworthy drivers, under current conditions, self-select into the anonymous regular taxi market, which still exists alongside the ride-sharing providers. In 2022, after the emergence of ride-sharing providers, we returned to Mexico City and collected new data in the same location and under a similar protocol as in our BASELINE treatment.[15] We hailed 32 cabs (no one refused this time). Again, we recorded a delivery rate of 100%, so there was no support for the selection-idea.

Unexpected forms of cheating

While there was virtually universal delivery, in some cases (seemingly not distinguishable by treatment), cab drivers cheated in other ways. Their gimmicks included telling RA-B that RA-A had not paid; showing up with a (possibly manipulated) meter read, vastly exceeding the amount originally agreed to and asking for a matching top-up; and claiming that some previously not mentioned extra fee applied. Due to the low number of instances, we lump the deviations from the agreement into two categories presented in Table 1: asking for extra money and asking to cover the cost. We find no statistical difference in the unexpected forms of cheating between any of the two treatments and two robustness checks. The results of statistical tests are available upon request.

5. Concluding remarks

While our results surprised us, in retrospect, and in light of another article by some of us, perhaps we should have known better? Dufwenberg et al. (2017) develop a theory of *informal agreements* in which one of two central assumptions is that once a person enters an informal agreement he or she will not renege.[16] When we designed our field experiment, we did not have that theory in mind. We were rather thinking in terms of a comparison to lab experiments with students, and

[13] We are not rejecting the predictions of Proposition 1. While its part (ii) can justify lower delivery rates in HAGGLING ($\delta > \varepsilon$), part (i) allows for universal delivery ($\delta = \varepsilon = 1$). See the proof in Appendix A.1 for exact conditions.

[14] The Mann-Whitney test p-value = 0.345 for initial offers; 0.526 for final offers; and 0.29 for the trip durations.

[15] We used a USB stick as an item to be delivered, instead of a CD (which in the meantime have mostly gone obsolete). Another difference is that the stated offers, even when adjusted for inflation, were significantly higher than those in the BASELINE (Mann-Whitney p-value < 0.001). This may be due to the fact that these days it takes longer time to get from point A to point B.

[16] The second assumption, less relevant for our purposes here, is that temptations to reneg affect the form of the informal agreements that people strike. For some other work on informal agreements, see also Miettinen (2013), Kessler and Leider (2012), Krupka et al. (2017), and Di Bartolomeo et al. (2023).

whether we could by treatment marginally affect trustworthiness in the directions described in the introduction. However, it appears that our design generates informal agreements that the shipped items will be delivered, and the 2017-theory would do a good job at explaining the data, even though our current experiment was never intended to test it.

That said, the data we have reported is, admittedly, and despite incorporating some robustness checks, limited as regards time, place, and details. It is premature to draw overly strong conclusions with confidence regarding how to understand the data. We would be happy if our study were merely given credit for being innovative with respect to introducing a new method for studying trustworthiness in the field. We are sympathetic with our referees who pointed to a plethora of aspects that may matter to the interpretation of our results, and that may warrant further scrutiny. Does it matter that trust games in the lab are not cast within a professional relationship whereas those we have created in the street are? How would taxi drivers behave in lab trust games? Can more be said about how cheating depends on the price, perhaps using treatments with greater variation in how prices are determined? Do the cab drivers really perceive the interaction as one-shot? Maybe one could run a survey with them to find out? How do cab drivers perceive the occurrence of haggling? Maybe taxi drivers do not perceive the negotiated price as unfair?[17] Or, could haggling be interpreted as a signal of the customer's competency or local knowledge, both of which could affect the likelihood of negative consequences for not delivering? Also, did the drivers believe the CD they were asked to deliver was a piracy piece? Maybe one should run treatments that vary the legality of the good? All of these questions were suggested to us by our referees. Although it is beyond the scope of our study to go down all these branches, in order to inspire future research, we invite our readers to contemplate various possibilities of getting at these follow-up research questions.

Declaration of competing interest

None.

Acknowledgments

We thank David Reiley, participants at several seminars, and two anonymous referees for helpful comments and discussion. Funding was provided by the University of Canterbury. We also thank Tamara Gonzáles-Ramírez, Roberto González-Téllez, and Hiram Serra-Peña for their excellent research assistance.

Appendix A

A.1. Price-haggling and reciprocity

Elements of D&K We focus mainly on C's utility which consists of a material component ($\pi_C(\cdot)$) and a reciprocity component. The latter is the product of how kind C believes that P is to him ($\lambda_{CPC}(\cdot)$) and how kind C is to P in return ($\kappa_{CP}(\cdot)$), weighted by C's reciprocity sensitivity parameter $Y_C \geq 0$. So C's utility has the form $\pi_C(\cdot) + Y_C \times \kappa_{CP}(\cdot) \times \lambda_{CPC}(\cdot)$.

In order to measure kindness, we need to consider the beliefs a player holds about the strategy of the other. A_i is i's set of behavioral strategies, $b_{ij} \in A_i$ is the first-order belief of i about j's strategy, and $c_{iji} \in A_i$ is i's second-order belief of about j's belief about i's strategy.[18] $\pi_j(a_i, b_{ij})$ is the (material) payoff i believes he gives to j (computed as if $a_j = b_{ij}$), and i is kind to j if i believes he gives j a relatively high payoff. Formally, i's kindness is computed by comparing $\pi_j(a_i, b_{ij})$ to the average, or "equitable," payoff that i believes that he could give j, given by

$$\pi_j^e(b_{ij}) = \frac{1}{2}\left(\max_{a_i \in A_i} \pi_j(a_i, b_{ij}) + \min_{a_i \in E} \pi_j(a_i, b_{ij})\right),$$

where $E_i \subseteq A_i$ contains those ("efficient") strategies of i that do not for sure lead to Pareto-inferior material outcomes in any history of play.[19] In our game $a_i \in A_i \backslash E_i$ iff $i = C$ and a_C puts positive probability on the choice *reject* following *counter*.[20]

i's kindness from choosing a_i when holding belief b_{ij}, is defined as

$$\kappa_{ij}\left(a_i, b_{ij}\right) = \pi_j(a_i, b_{ij}) - \pi_j^e(b_{ij})$$

and i's belief about the kindness of j, λ_{iji}, is derived the same way as κ_{ji}, replacing j's strategy a_j by b_{ij} and by replacing b_{ji} by c:

$$\lambda_{jij}\left(b_{ji}, c_{jij}\right) = \pi_j(b_{ji}, c_{jij}) - \pi_j^e(c_{jij}).$$

[17] A referee pointed out that "the literature finds a relatively weak positive relationship between employee prosociality and no-haggle pricing strategies (e.g., Kniffin et al., 2018)."

[18] All beliefs are point-beliefs, assigning probability 1 to whatever is believed.

[19] Strategies in $A_i \backslash E_i$ are called "inefficient" because they involve Pareto-decreasing "waste" after some history of play. Refer to D&K (pp. 275-7) for a precise definition and elaboration on why the $E_i \subseteq A_i$ feature is important to the theory.

[20] To see this, note that following choice *counter* choice *deliver* gives both P and C higher material payoff than choice *reject*.

In SRE, beliefs coincide with the chosen strategies and at every history of play beliefs are updated to be consistent with reaching that history. Furthermore, at all histories choices must be optimal given the beliefs.

Proof of Proposition 1. Recall that δ, ε, and ρ are the probabilities of, respectively, *deliver* following *accept*, *deliver* following *counter*, and *reject* following *counter*. We establish three lemmas around which the proof is built:

Lemma 1. In any SRE, $\rho = 0$. *To see this, note that if a_C puts a positive probability on* reject *following* counter, *then it is not an efficient strategy (as defined earlier in this section). It can never be rationally used except as an unkind response to a co-player believed to be unkind. However, in our game such a use can be ruled out, because if C believes that P is unkind, then it must be better for P to choose* not *(to deliver) than to* reject; *the former choice brings a higher payoff to C as well as a lower payoff to P than the latter choice does. This implies that* $\rho = 0$.

Lemma 2. In any SRE, $\delta \geq \varepsilon$. *Assume to the contrary that* $\delta < \varepsilon$. *P's kindness depends on his choices and on* b_{PC} *which specifies his beliefs about* δ, ε, *and* ρ. *In an SRE these beliefs are correct and it follows that P must be less kind following* counter *than following* accept. *To see this, refer to the definition of* $\kappa_{ij}(\cdot)$ *above (letting* $i = P$; $j = C$*) and note that since* $\rho = 0$ *(Lemma 1),* $p > p_r$, *and* $\delta < \varepsilon$ *we get*

$$\pi_C(\text{counter}, b_{PC}) = \varepsilon \cdot (p_r - t) + (1 - \varepsilon - \rho) \cdot p_r + \rho \cdot 0 = p_r - \varepsilon \cdot t$$
$$<$$
$$\pi_C(\text{accept}, b_{PC}) = \delta \cdot (p - t) + (1 - \delta) \cdot p = p - \delta \cdot t.$$

Moreover, since in SRE C's beliefs about P's beliefs are correct, C similarly perceives that P is more kind following accept *than following* counter. *Now note that the "marginal material impact" of C's choices, on himself as well as on P, is the same following* accept *as following* counter. *Namely, he can incur-or-not "t" for himself, and he can give or deny "*1*" to P). If* $1 > \varepsilon > \delta \geq 0$ *then C must be indifferent between* deliver *and* not *following* counter *(otherwise C wouldn't be willing to mix), but based on what we just said about the marginal material impact, it then follows that he must strictly prefer* deliver *to* not *following* accept. *Hence,* $\delta = 1$, *a contradiction. And if* $1 = \varepsilon > \delta \geq 0$, *then C must (weakly) prefer* deliver *to* not *following* counter, *but then again (based on what we said about the marginal material impact) it follows that he must strictly prefer* deliver *to* not *following* accept. *Hence,* $\delta = 1$, *again a contradiction. We conclude that* $\delta \geq \varepsilon$.

Lemma 3. In any SRE, if $0 < \varepsilon < 1$ then $\delta > \varepsilon$. *We verify this by contradiction. If the implication were false then either* $1 > \varepsilon = \delta > 0$ *or* $1 > \varepsilon > \delta \geq 0$ *would hold. Both of these cases can be considered simultaneously. Applying analogous arguments as in Lemma 2, we can conclude that* $\pi_C(\text{accept}, b_{PC}) > \pi_C(\text{counter}, b_{PC})$ *and that the marginal material impact of C's decision is the same following* accept *as following* counter. *Therefore, C must be indifferent between* deliver *and* not *in the subgame following* counter, *implying that* $\delta = 1$ *which is a contradiction.*

Part (i) of Proposition 1 is established by combining Lemmas 1 and 2.

It remains to prove part (ii) of Proposition 1. It is helpful to first establish necessary conditions on Y_C for the existence of SRE with, respectively, $\delta = \varepsilon = 0$ and $\delta = \varepsilon = 1$. For the former case, due to Lemma 2, we only need to check that $\delta = 0$ maximizes utility for C following *accept*. Plug relevant numbers into the utility expression $\pi_C(\cdot) + Y_C \times \kappa_{CP}(\cdot) \times \lambda_{CPC}(\cdot)$; one sees that $\kappa_{CP}(\cdot)$ equals $-\frac{1}{2}$ or $\frac{1}{2}$ depending on C's choice,[21] while $\lambda_{CPC}(\cdot)$ equals $p - \frac{1}{2} \cdot (p + 0) = \frac{p}{2}$, so the following inequality holds:

$$\underbrace{p + Y_C \cdot (-\frac{1}{2}) \cdot (\frac{p}{2})}_{\text{utility of } not} \geq \underbrace{p - t + Y_C \cdot (\frac{1}{2}) \cdot (\frac{p}{2})}_{\text{utility of } deliver}.$$

The inequality can be re-written as $Y_C \leq 2t$.

For the latter case ($\delta = \varepsilon = 1$), due to Lemma 2, we only need to check that $\varepsilon = 1$ maximizes utility for C following *counter*. Again $\kappa_{CP}(\cdot)$ equals $-\frac{1}{2}$ or $\frac{1}{2}$ depending on C's choice, but $\lambda_{CPC}(\cdot)$ now equals $p_r - t - \frac{1}{2} \cdot ((p - t) + 0) = \frac{2p_r - p - t}{2}$, so the following inequality holds:

$$\underbrace{p_r - t + Y_C \cdot (\frac{1}{2}) \cdot (\frac{2p_r - p - t}{2})}_{\text{utility of } deliver} \geq \underbrace{p_r + Y_C \cdot (-\frac{1}{2}) \cdot (\frac{2p_r - p - t}{2})}_{\text{utility of } not}.$$

[21] For *not* we get $\kappa_{CP} = 0 - \pi_C^e(b_{CP})$ and for *deliver* we get $\kappa_{CP} = 1 - \pi_C^e(b_{CP})$, where $\pi_C^e(b_{CP}) = \pi_C^e(accept) = \frac{1}{2}((1 - p) + (-p)) = \frac{1}{2}$.

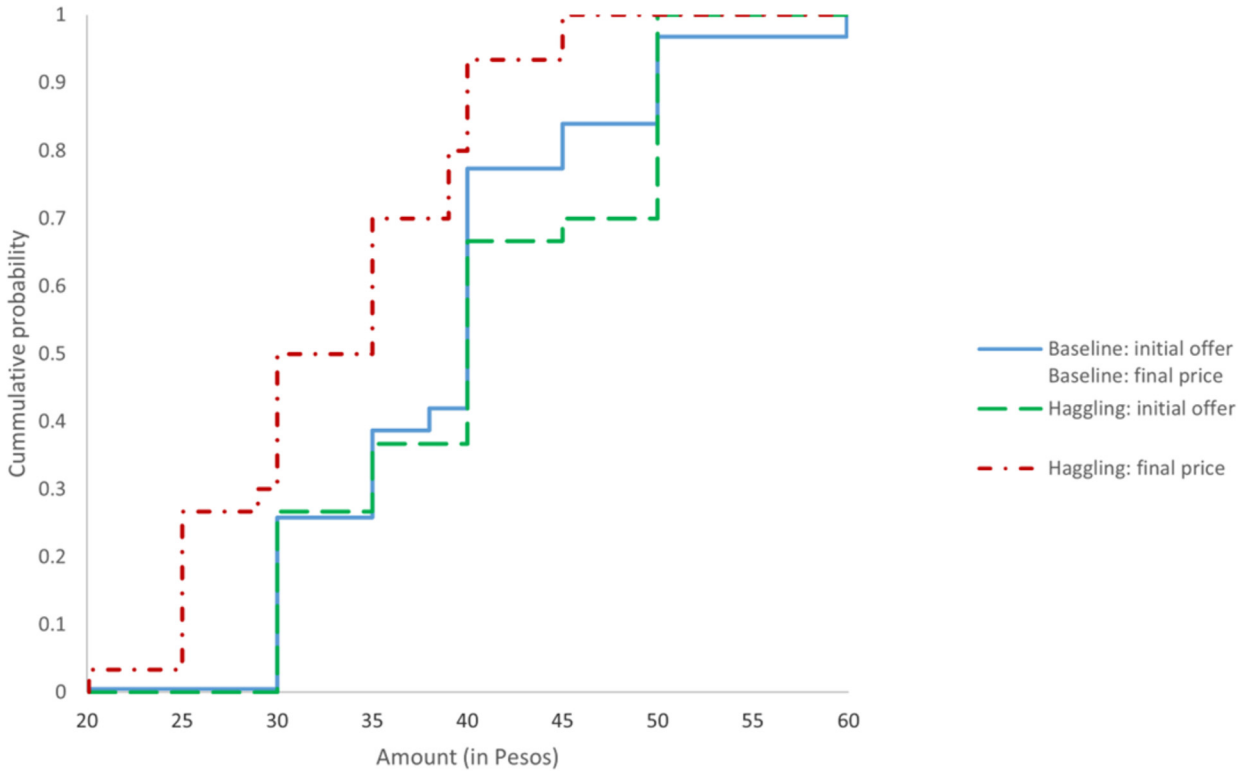


Fig. 1. Offer distributions: Baseline vs Haggling.

If $\frac{2p_r-p-t}{2} \leq 0$, then P's counteroffer is not interpreted as kind, and hence the inequality will not hold for any Y_C. If $\frac{2p_r-p-t}{2} > 0$, then we can re-write the inequality as $Y_C \geq \frac{2t}{2p_r-p-t}$. Note that $\frac{2t}{2p_r-p-t} > 2t$.

Now select ℓ and h, with $\ell < h$, such that neither of the necessary conditions for SRE with $\delta = \varepsilon = 0$ or $\delta = \varepsilon = 1$ hold: If $\frac{2p_r-p-t}{2} \leq 0$, select (any) $\ell > 2t$ and then $h > \ell$. If $\frac{2p_r-p-t}{2} > 0$, select $\ell \in (2t, \frac{2t}{2p_r-p-t})$ and then $h \in (\ell, \frac{2t}{2p_r-p-t})$. Suppose $Y_C \in (\ell, h)$. Some SRE must exist; this follows from D&K's existence theorem. By design of ℓ and h, neither $\delta = \varepsilon = 0$ or $\delta = \varepsilon = 1$ is compatible with SRE. By Lemma 2, $\delta \geq \varepsilon$. We can group the possibilities into three cases: $\varepsilon = 0 < \delta$ and $0 < \varepsilon \leq \delta < 1$ and $\varepsilon < \delta = 1$. For the first and last case, obviously $\varepsilon < \delta$. For the middle case, $\varepsilon < \delta$ is implied by Lemma 3. Part (ii) of Proposition 1 follows. □

A.2. Figures

Figs. 1–3 display the comparison of CDFs of fare offers for three pairs of treatments: 1. Baseline vs. Haggling, 2. Haggling vs. AltRoute (both initial offers and final offers), and 3. Baseline vs. Post-Uber. Fig. 4 shows the CDFs of trip-durations for all treatments.

A.3. Two routes

Our main route follows Vía Insurgentes (in Fig. 5, panel a), a major traffic artery that runs through some of the most affluent neighborhoods of Mexico City. Our secondary route (panel b) was similar in length and direction, but started downtown and did not follow a major commuter pathway. It also leads through some less wealthy, middle-class neighborhoods.

A.4. A survey

Would our findings be surprising also to locals? In a separate laboratory session conducted at ITAM, we invited 37 students, many of whom were Mexico City residents, to shed light on that question. We asked these subjects to guess the outcome of the behavior in seven different scenarios involving plausible trust situations around Mexico City; the second one was designed to be reminiscent of our experiment with the cab drivers:

1. You take a bus that is fully crammed with people. You manage to jump on in the back door. The only way for you to pay the fare is to send the money up front to the driver by asking passengers to pass it on up front. The bus fare is 5 pesos but you do not have any change so your only option is to pass 20 pesos. Will you get the change back?

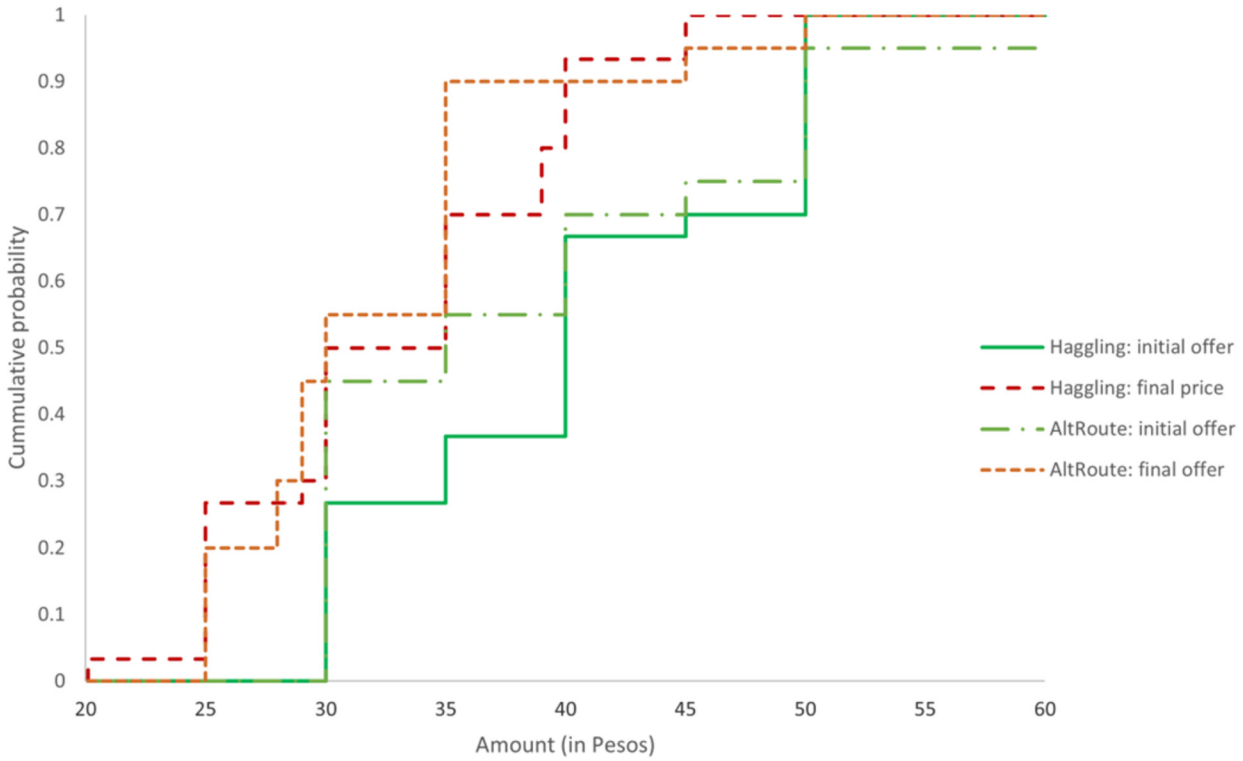


Fig. 2. Offer distributions: Haggling vs AltRoute.

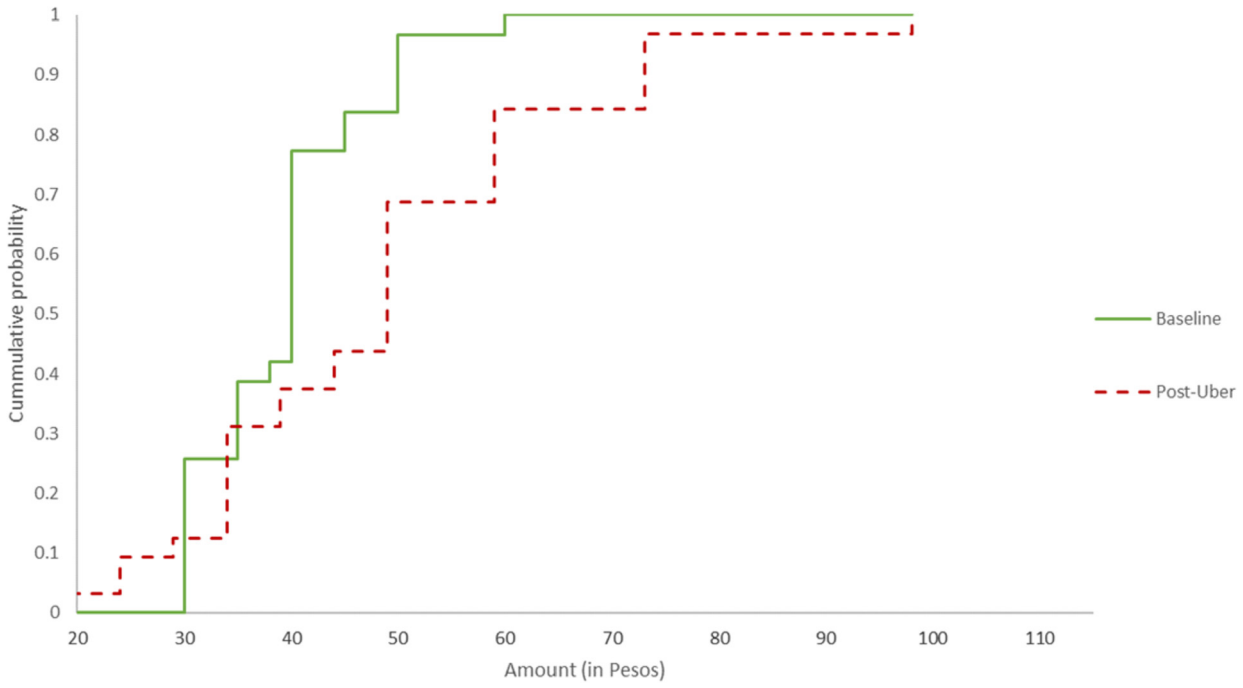


Fig. 3. Offer distributions: Baseline vs Post-Uber.

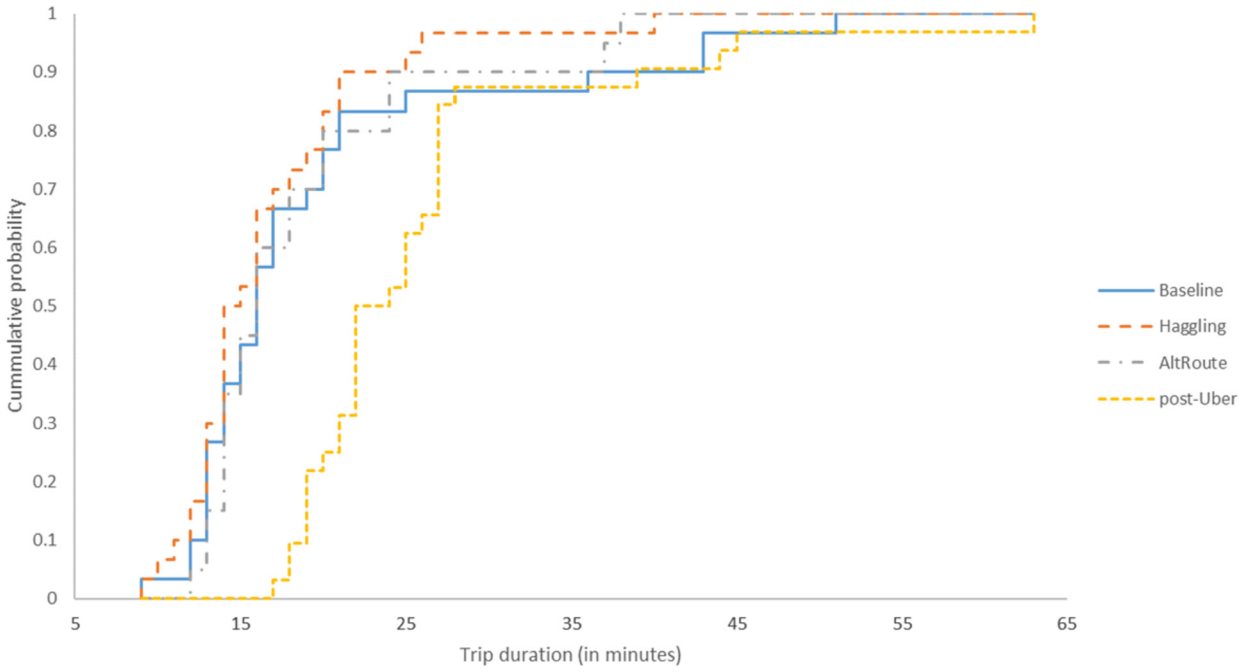


Fig. 4. Trip durations.

(a) Via Insurgentes (b) Calle Bolivar

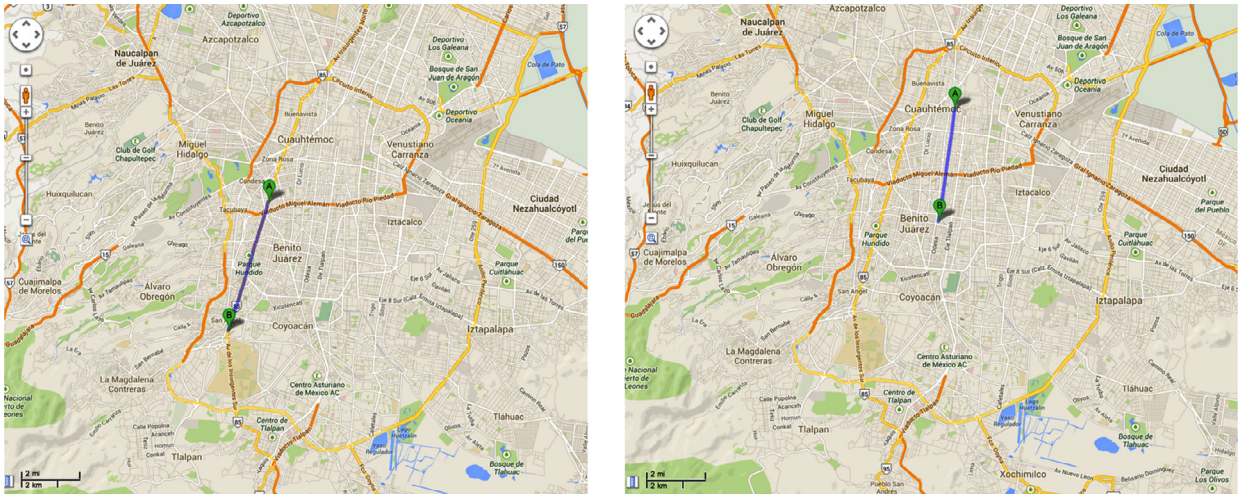


Fig. 5. Routes.

2. You need to pass a school project to your friend which is due the next day. The project contains sounds and video clips so you burn it on a CD. You cannot leave the house because you have to watch after your little brother. If you flag a cab on the street and pay the amount he asks ahead of time will the CD get to your friend?
3. You are at a street market (mercado sobre ruedas) looking for a gift for your friend. You come across a seller who is selling artisan tiles with custom writings on them. You would like to have one of those made with your friend's name on it. The seller wants the full payment of 100 pesos up front and promises to bring you the tile (same place same time) one week later. If you decide to go ahead and pay him, will you get your tile one week later?
4. You are at a football match and have to use a restroom. There is no assigned seating and you happen to have a good seat. If you leave your jacket on your seat as a place holder will it still be there after you've come back?
5. You are in a bar around Centro Historico with a group of five friends celebrating your birthday. You feel generous and offer to pick up the tab for the night. If you give the credit card to the bartender will the final tab at the end of the night be correct?

Table 2
Frequency of trusting responses.

	Scenarios						
	1	2*	3	4	5	6	7
All data:							
Freq.	30	21	25	14	12	3	20
%	81	57	68	38	32	8	54
Mex. City natives:							
Freq.	21	13	20	11	10	3	12
%	78	48	74	41	37	11	44

Note: * denotes our cab scenario.

6. You go to the stadium to buy tickets for a football match that will take place tomorrow but they are sold out. A man outside the ticket office (a scalpel) offers to get you tickets at 30% discount. He does not have the tickets on him but has to walk over to his friend to get them. If you give him the money will he show up with the tickets?
7. You are going to a birthday party in La Condesa but can't find a spot to park your car. You have just come from a long road-trip and your car is full of personal belongings bags etc. in the back and front seat. If you leave the car at a Valet Parking will everything be there when you pick it up?

Responses were incentivized using the Houser and Xiao (2011) payment procedure according to which subjects get paid if their answers match the answer of another randomly selected participant in the room. This effectively creates a coordination game among the subjects. Our thinking was that if a subject thought it obvious that the correct answer would be yes, then they would attempt to coordinate on the corresponding equilibrium in the procedure. Overall, such coordination did not happen. In the second scenario, 43% of subjects guessed that the cab would not deliver; the rest guessed that the cab would deliver. Among those born in Mexico City, 52% thought the cab would not deliver. This suggests that it is not common knowledge among locals that the correct answer should be yes. The frequencies of trusting responses for all scenarios for full vs. restricted-to-Mexico-City-subjects sample are reported in Table 2.[22]

References

Abeler, J., Nosenzo, D., Raymond, C., 2019. Preferences for truth-telling. Econometrica 87 (4), 1115–1153.
Azar, O.H., Yosef, S., Bar-Eli, M., 2013. Do customers return excessive change in a restaurant? A field experiment on dishonesty. J. Econ. Behav. Organ. 93, 219–226.
Balafoutas, L., Nikiforakis, N., 2012. Norm enforcement in the city: a natural field experiment. Eur. Econ. Rev. 56 (8), 1773–1785.
Balafoutas, L., Beck, A., Kerchsbamer, R., Sutter, M., 2013. What drives taxi drivers? A field experiment on fraud in a market for credence goods. Rev. Econ. Stud. 80, 876–891.
Balafoutas, L., Kerchsbamer, R., Sutter, M., 2017. Second-degree moral hazard in a real-world credence goods market. Econ. J. 127, 1–18.
Battigalli, P., Dufwenberg, M., 2022. Belief-dependent motivations and psychological game theory. J. Econ. Lit. 60, 833–882.
Bengtsson, N., 2016. Efficient informal trade: theory and evidence from the Cape Town taxi market. J. Dev. Econ. 115, 85–98.
Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142.
Bhattacharya, H., Dugar, S., 2020. The hidden cost of bargaining: evidence from a cheating-prone marketplace. Int. Econ. Rev. 61 (3), 1253–1280.
Castillo, M., Petrie, R., Torero, M., Vesterlund, L., 2013. Gender differences in bargaining outcomes: a field experiment on discrimination. J. Public Econ. 99, 35–482.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 117, 817–869.
Cohn, A., Maréchal, M.A., Tannenbaum, D., Zünd, C.L., 2019. Civic honesty around the globe. Science 365 (6448), 70–73.
Di Bartolomeo, G., Dufwenberg, M., Papa, S., Passarelli, F., 2023. Promises or agreements? Moral commitments in bilateral communication. Econ. Lett. 222, 110931.
Dreber, A., Johannesson, M., 2008. Gender differences in deception. Econ. Lett. 99 (1), 197–199.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47, 268–298.
Dufwenberg, M., Servátka, M., Vadovič, R., 2017. Honesty and informal agreements. Games Econ. Behav. 102, 269–285.
Dugar, S., Bhattacharya, H., 2017. Fishy behavior: a field experiment on (dis) honesty in the marketplace. J. Behav. Exp. Econ. 67, 41–55.
Dugar, S., Mitra, A., Shahriar, Q., 2019. Deception: the role of uncertain consequences. Eur. Econ. Rev. 114, 1–18.
Dulleck, U., Kerschbamer, R., Sutter, M., 2011. The economics of credence goods: an experiment on the role of liability, verifiability, reputation, and competition. Am. Econ. Rev. 101 (2), 526–555.
Ellingsen, T., Johannesson, M., Lilja, J., Zetterqvist, H., 2009. Trust and truth. Econ. J. 119 (534), 252–276.
Falk, A., Zehnder, C., 2013. A city-wide experiment on trust discrimination. J. Public Econ. 100 (C), 15–27.
Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. J. Econ. Perspect. 14, 159–181.
Finseraas, H., Hanson, T., Johnsen, Å., Kotsadam, A., Torsvik, G., 2019. Trust, ethnic diversity, and personal contact: a field experiment. J. Public Econ. 173, 72–84.
Fischbacher, U., Föllmi-Heusi, F., 2013. Lies in disguise – an experimental study on cheating. J. Eur. Econ. Assoc. 11 (3), 525–547.
Gibson, R., Tanner, C., Wagner, A.F., 2013. Preferences for truthfulness: heterogeneity among and within individuals. Am. Econ. Rev. 103 (1), 532–548.
Gneezy, U., 2005. Deception: the role of consequences. Am. Econ. Rev. 95 (1), 384–394.

[22] In the experiment, we used two different presentation orderings: Ord-1 and Ord-2. In Ord-1 our cab scenario was listed second in the sequence and in Ord-2 it was listed sixth. Ord-1 was run with 18 subjects and Ord-2 with 19 subjects. There were no significant differences in responses between the two orderings which is why we pooled the data.

Gneezy, U., Kajackaite, A., Sobel, J., 2018. Lying aversion and the size of the lie. Am. Econ. Rev. 108 (2), 419–453.
Gottschalk, F., Mimra, W., Waibel, C., 2020. Health services as credence goods: a field experiment. Econ. J. 130 (629), 1346–1383.
Houser, D., Xiao, E., 2011. Classification of natural language messages using a coordination game. Exp. Econ. 14, 1–14.
Houser, D., Vetter, S., Winter, J., 2012. Fairness and cheating. Eur. Econ. Rev. 56 (8), 1645–1655.
Huck, S., Konrad, K.A., Müller, W., Normann, H.T., 2007. The merger paradox and why aspiration levels let it fail in the laboratory. Econ. J. 117 (522), 1073–1095.
Huck, S., Lünser, G.K., Tyran, J.R., 2012. Competition fosters trust. Games Econ. Behav. 76 (1), 195–209.
Innes, R., Mitra, A., 2013. Is dishonesty contagious? Econ. Inq. 51 (1), 722–734.
Kerschbamer, R., Sutter, M., 2017. The economics of credence goods: a survey of recent lab and field experiments. CESifo Econ. Stud. 63 (1), 1–23.
Kessler, J.B., Leider, S., 2012. Norms and contracting. Manag. Sci. 58 (1), 62–77.
Kniffin, K.M., Reeves-Ellington, R., Wilson, D.S., 2018. When everyone wins? Exploring employee and customer preferences for no-haggle pricing. Front. Psychol. 9, 1555.
Krupka, E.L., Leider, S., Jiang, M., 2017. A meeting of the minds: informal agreements and social norms. Manag. Sci. 63 (6), 1708–1729.
Lundquist, T., Ellingsen, T., Gribbe, E., Johannesson, M., 2009. The aversion to lying. J. Econ. Behav. Organ. 70 (1–2), 81–92.
Miettinen, T., 2013. Promises and conventions – an approach to pre-play agreements. Games Econ. Behav. 80, 68–84.
Pruckner, G.J., Sausgruber, R., 2013. Honesty on the streets: a field study on newspaper purchasing. J. Eur. Econ. Assoc. 11 (3), 661–679.
Rabin, M., 1993. Incorporating fairness into game theory and economics. Am. Econ. Rev. 83, 1281–1302.
Schneider, H.S., 2012. Agency problems and reputation in expert services: evidence from auto repair. J. Ind. Econ. 60 (3), 406–433.
Sobel, J., 2005. Interdependent preferences and reciprocity. J. Econ. Lit. 43, 396–440.
Sutter, M., 2009. Deception through telling the truth?! Experimental evidence from individuals and teams. Econ. J. 119 (534), 47–60.
Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. Econometrica 76, 1467–1480.