



Building trust: The costs and benefits of gradualism[☆]

Melis Kartal^a, Wieland Müller^{b,c,*}, James Tremewan^d

^a Vienna University of Economics and Business & VCEE, Austria

^b University of Vienna & VCEE, Austria

^c Tilburg University, CentER & TILEC, Netherlands

^d University of Auckland, New Zealand



ARTICLE INFO

Article history:

Received 5 March 2020

Available online 23 August 2021

JEL classification:

C73

C91

C92

D82

D83

Keywords:

Cooperation

Trust

Infinitely repeated games

Gradualism

Game theory

Experiments

ABSTRACT

We examine the prevalence of gradualist strategies and their effect on trust-building and economic gains in a setting with an infinite horizon, asymmetric information regarding the trustworthiness of receivers, and various levels of trust. The theoretical literature suggests that gradualist strategies mitigate asymmetric information problems and foster trust-building. However, we theoretically and experimentally show that gradualism sometimes reduces joint payoffs relative to a simple “binary” setting in which trust is an all-or-nothing decision. In a series of experiments, we vary the degree of asymmetric information as well as the economic returns to trusting behavior, and delineate circumstances under which gradualism may promote or curb efficiency.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Many economic relationships lack formal enforcement and rely on trust for generating mutual gains from trade.¹ One major concern at the beginning of such relationships is the prospect of facing an opportunistic partner who is around for short term gains. This is why a first-time buyer may be concerned about the credibility of a supplier, a lender about the repayment incentives of a first-time borrower, and a manager about the tendency of a new employee to shirk responsibilities. The literature suggests that under such asymmetric information, *gradualism* is a natural candidate for equilibrium trust profiles (i.e., the relationship starts small with testing levels of trade, and stakes in the relationship increase after successful

[☆] We are grateful to an anonymous Associate Editor and three anonymous referees for very helpful comments. We also thank Guillaume Fréchette, Steffen Huck, Jens Prüfer, Karl Schlag, Martin Salm, Andy Schotter and members of seminar audiences at the Universities of Konstanz, Tilburg, Nottingham, Birmingham, Lyon and Copenhagen, the University of Economics Prague, the Higher School of Economics Moscow, the Berlin IO Day, the International Symposium in Experimental Economics at NYU Abu Dhabi, and Workshop “The Neoclassical Repair Shop and its Alternatives” at WZB Berlin for helpful comments. The research reported in this paper is funded by the University of Vienna.

* Corresponding author.

E-mail address: wieland.mueller@univie.ac.at (W. Müller).

¹ This is the subject matter of a large literature on self-enforcing agreements and relational contracts (see, among others, Bull (1987), MacLeod and Malcolmson (1989), Baker et al. (2002), and Levin (2003)).

trades).² The positive effects of gradualism seem to be confirmed by its prevalence in practice, for example, in markets (see, e.g., McMillan and Woodruff, 1999; Rauch and Watson, 2003) and microcredit lending (Morduch, 1999). In the latter context, this practice is called “progressive lending” and involves gradually increasing loan sizes for borrowers who are in good standing.³ Of course, the role of gradualism in relationships is not limited to market settings. For example, many marriages are initiated after partners go through numerous smaller steps, such as dates, vacations, cohabitation, and engagement.

Gradualist strategies can—in theory—circumvent asymmetric information problems and foster trust-building by limiting the payoff loss from a betrayal. However, gradualist strategies are quite sophisticated in that people may easily “test” too little or too much relative to the optimal level. Furthermore, it is not clear gradualism maximizes overall efficiency, for example, relative to simpler settings where options for “starting small” are limited. Therefore, we investigate the efficiency properties of gradualism not only within but also across games. While theory so far has emphasized the potential benefits of gradualism, our framework also highlights the possible costs. Importantly, we theoretically and experimentally delineate circumstances under which gradualism performs better or worse than a natural benchmark that restricts the scope for testing strategies.

We consider settings in which two players, a sender and a receiver, play a generalized version of the so-called trust game infinitely many times. The sender moves first by choosing either no trust or one of three positive trust levels: low trust, medium trust or high trust. If a positive trust level is chosen, the receiver chooses whether to return or default, whereas if the sender chooses no trust, the receiver has no choice to make, and both parties receive their outside options. We call such a game the “gradual game.” The receiver has private information regarding his type, which is either high or low. The high-type receiver is a cooperative commitment type who always returns trust, whereas the low type is in principle untrustworthy; that is, for the low type it is strictly optimal to (eventually) default given any sender strategy involving trust.

All equilibria of the gradual games in our experiment are characterized by gradualism: the sender starts by choosing low trust and increases the trust level gradually as long as the receiver returns. Because the payoff loss for the sender from receiver default is smaller at lower trust levels, gradualist strategies make senders better off by limiting potential default losses. In the experiment, we measure the extent to which sender behavior in the gradual game is consistent with our game-theoretic predictions.

For every gradual game, we introduce a corresponding “binary game”: this game differs from its gradual version only in that low and medium trust levels are absent—all other parameters are held constant. In the gradual game, the sender uses “gradualist incentives” in addition to “dynamic incentives” (i.e., static Nash reversion as punishment for default). In the binary game, the sender can only use dynamic incentives.

While the previous literature emphasized that gradualism promotes trust building and efficiency, neither theory nor experimental work considered the efficiency gains from gradualist strategies against a restricted reference, such as the binary game. In theory, whether gradualism always improves joint efficiency is not obvious. Therefore, we develop a theoretical framework and outline conditions under which the gradual game performs better or worse than its binary version. In our framework, the sender is always (weakly) better off in the gradual game than in the binary game, but this is not true for the receiver nor for joint efficiency. To see why, first note that if the prior probability of meeting a high-type receiver is relatively low, then the sender would take the limited risk of testing the receiver in the gradual game, but avoid the higher risk of trusting in the binary game. In this case, both the sender and receiver are better off in the gradual game.⁴ As the prior probability of meeting a high type increases, there will be a range of priors such that (i) the sender still strictly prefers a gradualist strategy in the gradual game as this limits the risk exposure of the sender, and yet, (ii) the sender will also take a chance on the receiver and trust in the binary game as the risk of default is not too high. For this range of priors, the receiver is strictly better off in the binary game than in the gradual game, since the higher the trust level, the higher the payoff of the receiver. In fact, we show that joint payoffs will also be strictly higher in the binary game than in the gradual game for a subset of this range of priors. In line with our discussion above, whether or not the gradual game performs better is characterized by a threshold prior probability of meeting a high-type receiver. That is, if the prior probability of meeting a high type is below the threshold, then the gradual game makes both senders and receivers better off relative to its binary counterpart, but if the prior exceeds the threshold, the gradual game is no longer advantageous, and joint efficiency may be strictly higher in the binary game. The threshold probability can be interpreted as a statistic reflecting the sender’s risk when trusting in the binary game: the higher the threshold, the higher the risk. The threshold probability depends on the game parameters, which we vary across our experimental games.

We have in total four pairs of gradual and binary game treatments. In these treatment pairs, we manipulate sender payoffs as well as the probability of meeting a high type using different types of computerized receivers in order to test our theoretical predictions. Experimental results provide support for the gradualism hypothesis. A large majority of the sender behavior in our four gradual games can be classified as belonging to one of six intuitive categories, and the category

² See, e.g., Sobel (1985), Ghosh and Ray (1996), Kranton (1996), Watson (1999, 2002), Halac (2012), and Kartal (2018).

³ Morduch (1999, p. 1582) writes: “[Microcredit] Programs typically begin by lending just small amounts and then increasing loan size upon satisfactory repayment. The repeated nature of the interactions—and the credible threat to cut off any future lending when loans are not repaid—can be exploited to overcome information problems and improve efficiency ...”.

⁴ More generally, the gradual game always weakly improves the equilibrium payoff of the sender as any sender strategy in the binary game can be identically implemented in the gradual game, and using testing strategies may strictly increase the sender payoff in the gradual game.

comprising behavior consistent with equilibrium predictions is statistically the largest. Furthermore, given empirical receiver behavior, equilibrium strategies make senders better off than non-equilibrium strategies as predicted.

Regarding the efficiency effect of gradualism, we observe that the experimental results are overall consistent with our theoretical framework. The gradual game performs better than its binary version in terms of joint payoffs *only* in the treatment pair where there are computerized receivers programmed to always default, and the payoff return to the sender from trust is relatively low. Under these two conditions, taking a chance on the receiver is a risky endeavor in the binary game, and thus, many senders are unwilling to trust, whereas in the gradual game senders frequently use testing strategies, which is conducive to trust and boosts cooperation. In the remaining treatment pairs, the sender's risk when trusting in the binary game is significantly lower, and in line with this, most senders choose to trust in the binary game. As a result, the presence of gradualist strategies either does not benefit or even hampers efficiency in these treatment pairs.

Our theoretical and experimental findings suggest that the merits of gradualism as a trust-building mechanism underscored in the previous literature are conditional. In fact, restricting the scope for gradualism may boost joint efficiency if a sufficiently large proportion of receivers is high type implying that the sender faces only a moderate risk of default in the binary game. This may have important practical implications. As an example, progressive lending by microcredit institutions is strictly gradualist and relies on lending only very small amounts at the beginning. It is usually argued that initial loans are too small to help borrowers. Therefore, whether such steep progressivity in the lending schedule is too prudent or efficient is a relevant empirical question especially for non-profit lenders, which care about borrower welfare. Field research varying the progressivity of the lending schedule could inform policy makers and non-profit lenders on how to efficiently tailor the loan size and growth over time. This is a possible direction for future research.

2. Related literature

There is an extensive literature on reputation and cooperation building that dates back to the seminal works by Kreps and Wilson (1982), Milgrom and Roberts (1982) and Kreps et al. (1982). While Kreps and Wilson (1982) and Milgrom and Roberts (1982) analyze an entry-deterrence game, Kreps et al. (1982) show that incomplete information about players' preferences can sustain cooperation in the equilibrium of a finitely repeated prisoners' dilemma game.⁵

Experimenters have implemented the model by Kreps et al. (1982) and its variants in the laboratory in order to test the sequential equilibrium predictions. While Andreoni and Miller (1993) study a finitely repeated prisoners' dilemma game as in Kreps et al. (1982), Camerer and Weigelt (1988), and the subsequent studies by Neral and Ochs (1992), Anderhub et al. (2002), Brandts and Figueras (2003), and Grosskopf and Sarin (2010) study whether sequential equilibrium is successful in predicting behavior in a finitely repeated binary trust game, varying different dimensions in the design (e.g., the probability with which the receiver is a trustworthy type, the observability of past choices of the receiver to the sender, or whether the trustworthy type is computerized). Brown et al. (2004) consider finitely repeated gift-exchange games between employers and workers and show that initial behavior of both the employer and the worker is important for the functioning of successful "long-term" relations with high-wage and high-effort choices. Brown and Serra-Garcia (2017) analyze a finitely repeated lending game in which a creditor can make one of several loan offers to a borrower to study the effect the threat of terminating the relationship has on the nature and functioning of implicit loan contracts. Cocharde et al. (2004) study a finitely repeated investment game to analyze the explanatory power of reputation building as well as reciprocity with respect to their experimental data. Huck et al. (2016) use a binary trust game to study reputation building for quality on the part of a seller.

The experimental papers discussed above study games that are finitely repeated. In the equilibrium of a finitely repeated game, it is not generally possible to build up a relationship gradually because in every new period, there are fewer periods remaining, and the future value of the relationship is constantly decreasing, which reduces the incentive of the sender to increase stakes gradually.⁶ This is also empirically validated as Brown et al. (2004) find that successful relationships are *not* gradualist, but rather exhibit high wage and high effort from the very beginning. Furthermore, many economic relationships realistically have an indefinite time span rather than a fixed known horizon. Therefore, we focus on infinitely repeated trust games with variable trust levels. We show that equilibrium strategies are gradualist, and these strategies enable the sender and the high type receiver to reach the efficient level of trade after a testing period. This connects our paper to a branch of the theoretical reputation literature that draws a similar conclusion in diverse settings (see, e.g., Sobel, 1985; Ghosh and Ray, 1996; Kranton, 1996; Watson, 1999, 2002; Halac, 2012; Kartal, 2018). In addition, our paper relates to a burgeoning experimental literature on infinitely repeated games.⁷ Most of this experimental literature has so far focused on simple stage games—for exceptions, see e.g., Vespa and Wilson (2019), Wilson and Vespa (2020), and Bernard et al. (2018). Bernard et al. (2018) is also related as they study an infinitely repeated labor market setting with various possible cooperation levels and show that their experimental findings can be organized by a simple model where all employees are one of two behavioral

⁵ If each player is a type that is "committed" to a *tit-for-tat* strategy with some positive probability, then reputation concerns can motivate self-interested players to imitate a committed type for some time and generate cooperative behavior in a finitely repeated prisoners' dilemma game.

⁶ See as an example the theoretical predictions in Camerer and Weigelt (1988). The sender is less likely to choose trust over the course of the game even if the receiver has always returned in the past and holds a perfect record.

⁷ See, e.g., Palfrey and Rosenthal, 1994; Engle-Warnick and Slonim, 2004, 2006a, 2006b; Dal Bó, 2005; Aoyagi and Fréchette, 2009; Camera and Casari, 2009; Duffy and Ochs, 2009; Dal Bó and Fréchette, 2011, 2019; Fudenberg et al., 2012.

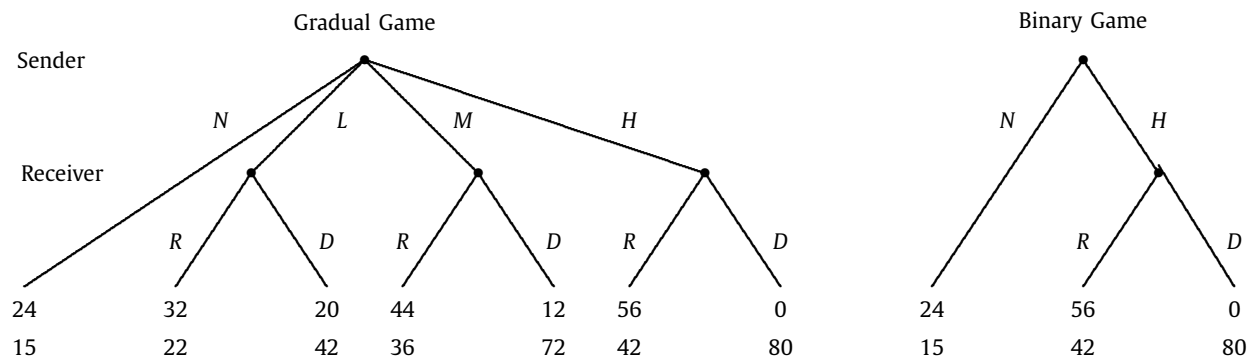


Fig. 1. The game trees of the baseline gradual and binary games.

types: cooperative or uncooperative. Our design enables us to have very precise theoretical predictions to measure the extent of equilibrium gradualist behavior, and we evaluate theoretically and experimentally the effectiveness of gradualism on fostering trust. Unlike the previous theoretical and experimental literature, our study highlights not only the benefit but also the potential cost of gradualism relative to a simpler setting that restricts the scope for gradualist strategies.

3. Model

We analyze an infinitely repeated game with the following features. At the beginning of each period $t = 0, 1, \dots$, the sender chooses either no trust (N) or one of $n \geq 2$ positive trust levels. If the sender chooses a positive trust level, then the receiver chooses between return (R) and default (D). In order to reduce notation and conserve space in the text, we assume that $n = 3$ as in our experimental design. That is, at the beginning of each period, the sender chooses either no trust (N) or one of three positive trust levels: low trust (L), medium trust (M), and high trust (H). In Online Appendix A.2, we present and prove the main theoretical result of this section in the general setting with arbitrary n and payoffs. The respective monetary payoffs for the sender and the receiver in the stage game are consistent with the typical trust game à la Berg et al. (1995):

- The sender and the receiver have conflicting interests. If a positive trust level is chosen, then the sender is strictly better off if the receiver returns, whereas the receiver is strictly better off if he defaults. In addition, the higher the trust level, the lower (higher) the payoff of the sender (receiver) if the receiver defaults.
- The payoff from mutual cooperation (i.e., choice of a positive trust level followed by return) strictly increases in the trust level for both players.

We call this game with multiple levels of trust the “gradual game” and denote it by Γ_G . Next, Γ_B denotes the “binary game” version of Γ_G in which sender choices are restricted to only N and H . Fig. 1 presents the gradual and binary game payoffs in our baseline treatments as an example. In both Γ_G and Γ_B , stage game payoffs for the sender and the receiver are discounted in future periods according to δ_S and δ_R , respectively.

The receiver is privately informed about his type, which is either high or low. The sender only knows the prior probability that the receiver is a high type, denoted by μ . The high type receiver is a cooperative type who always returns the sender's trust regardless of payoffs and δ_R . The low-type receiver maximizes his monetary payoff and is *untrustworthy* in the context of our games as follows: the value of δ_R and the stage-game payoffs are such that for *any* trust profile, the low-type receiver will—eventually—default.⁸ Since the high type never defaults, hereafter we focus only on the behavior of the sender and the low-type receiver, and *receiver* refers to the low type unless otherwise stated.

We compare Γ_G to its binary version Γ_B theoretically and experimentally regarding their trust-fostering and efficiency-enhancing properties under asymmetric information. Previous theoretical studies suggest that gradualism is a natural candidate for equilibrium profiles in Γ_G and highlight it as a successful trust-building mechanism. While the restricted game Γ_B does not allow for gradualism, the sender in Γ_B may still discipline the receiver via the threat of static Nash reversion. Therefore, Γ_B is a simple but relevant reference with which the potential benefit of gradualism in Γ_G can be compared. Although the sender is always (weakly) better off in Γ_G than in Γ_B , whether relationships are unambiguously more efficient in Γ_G than in Γ_B is not clear.

The magnitude of the prior μ is of utmost importance regarding the comparison of Γ_G to its restricted version Γ_B . To show this, let μ_B denote the value of μ that makes the sender in Γ_B exactly indifferent between the non-trusting strategy “always N ” and the conditionally trusting strategy that starts with H .⁹ In every gradual game of our experiment, sender

⁸ As will become clear below, the low-type receiver may optimally default at a later round in Γ_G if the sender uses a gradualist strategy.

⁹ More formally, the conditionally trusting strategy is such that the sender starts by choosing H at $t = 0$ and chooses H if the receiver returned in all previous periods, and always N otherwise.

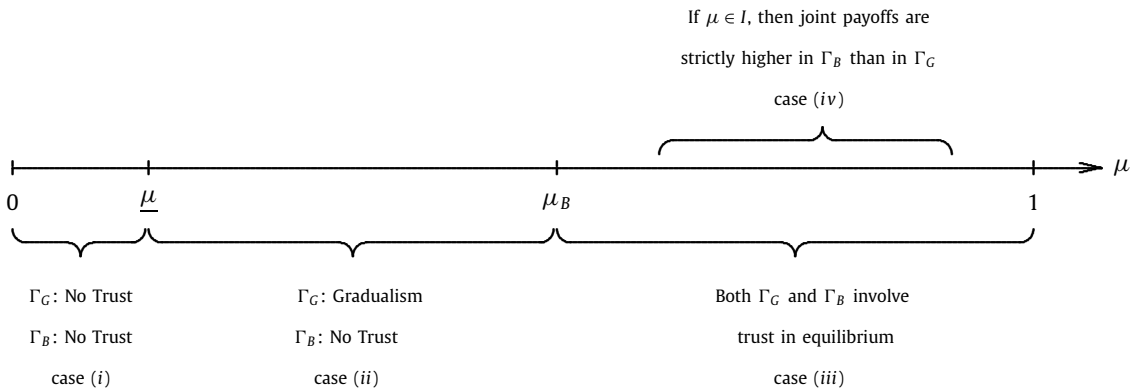


Fig. 2. Illustrating the possible cases of Proposition 1.

payoffs are such that if $\mu = \mu_B$, then both “always N ” and the conditionally trusting strategy that starts with H are strictly dominated by a *gradualist* strategy for the sender in Γ_G . This is a sufficient condition to prove the following result for our experimental gradual games and their binary versions (we present and prove the general result in Online Appendix A.2).

- Proposition 1. (i)** *There exists $\underline{\mu} \geq 0$ such that for $\mu < \underline{\mu}$, neither Γ_B nor Γ_G involves trust in equilibrium.*
(ii) Benefit of gradualism: *For all $\mu \in [\underline{\mu}, \mu_B)$, both the sender and the receiver are better off in Γ_G than in Γ_B .*
(iii) *For all $\mu \geq \mu_B$, the sender is better off, and the receiver is worse off in Γ_G than in Γ_B (the effect on joint expected payoffs is ambiguous).¹⁰*
(iv) Cost of gradualism: *There exists a non-empty, non-singleton interval $I \subset [\mu_B, 1)$ such that for all $\mu \in I$, joint expected payoffs are strictly higher in Γ_B than in Γ_G .*

The intervals of priors corresponding to parts (i)-(iv) of Proposition 1 are illustrated in Fig. 2. Part (i) of Proposition states that with very low levels of the prior, even gradualist strategies are unattractive to senders, and there will be no trust in either type of game. Part (ii) of Proposition 1 highlights the advantage of gradualism consistent with the previous theoretical literature: the gradual game Γ_G can foster trust for priors that are lower than μ_B unlike its binary version Γ_B . That is, if the prior probability of meeting a high-type receiver is low (but not too low so that $\mu \geq \underline{\mu}$), then the sender would only trust when the default risk can be mitigated by use of gradualist strategies, which are not available in the binary game. In other words, for this range of priors the gradual game equilibrium involves trust-building in equilibrium, whereas trusting the receiver is too risky in the binary version – here the ability to use gradualist strategies increases efficiency. Once the prior probability of meeting a high type is high enough and exceeds μ_B , the gradual game loses its trust-fostering advantage over the binary game (part (iii) of Proposition 1). While the sender still prefers a gradualist strategy in the gradual game for most of this range of priors, the sender will also take a chance on the receiver and risk trusting in the binary game as the share of high type receivers is high enough. More formally, if $\mu \geq \mu_B$, then the conditionally trusting strategy that starts with H is optimal in Γ_B . Moreover, in this range of priors the preference of senders for gradualist strategies can limit joint payoffs relative to those in the binary game, since lower levels of trust are achieved during the phase of trust-building (part (iv) of Proposition 1).

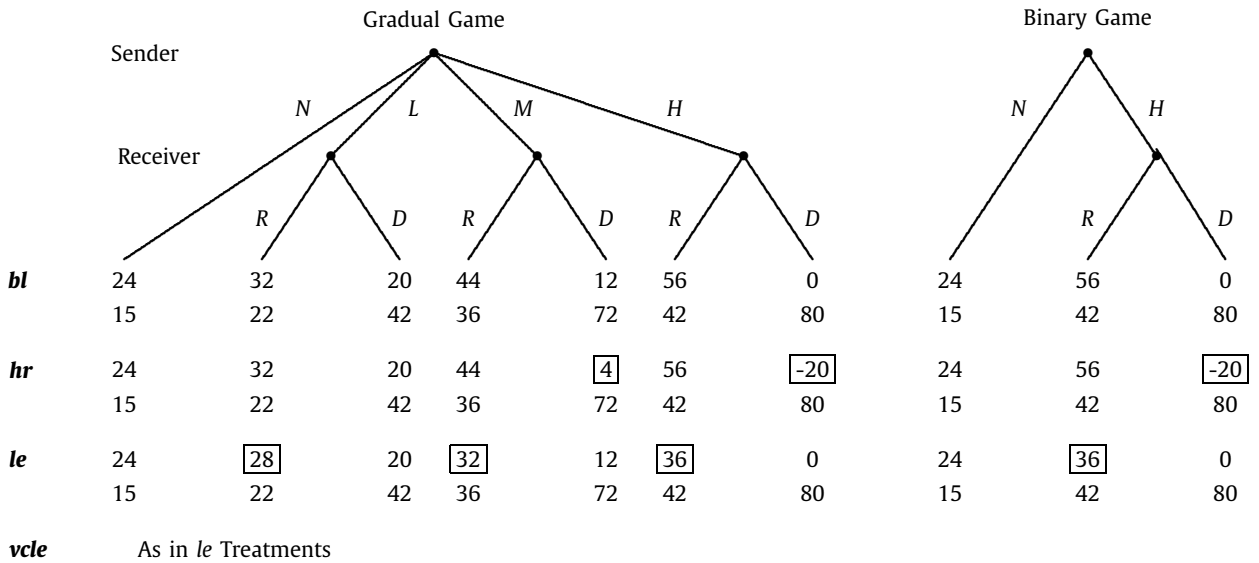
As a result, whether the prior probability μ is below or above μ_B in practice is crucial to understand the efficiency impact of gradualism relative to the restricted benchmark Γ_B . Our experimental design presents gradual and binary games that systematically vary μ and μ_B to test the predictions of Proposition 1 and identify conditions under which the presence of gradualist strategies generates costs or benefits.

4. Experimental design and predictions

The experiment consists of several infinitely-repeated gradual and binary game treatments. In all treatments, $\delta_S = 0.75$ and $\delta_R = 0.5$. At the beginning of each session, each subject is randomly assigned the role of a sender or a receiver. Subjects remain in the same role throughout the session. Every treatment involves eight senders and eight receivers. In every treatment, one of the eight receivers is a *computerized* player that is programmed to always return (i.e., a *high-type* receiver).¹¹ The probability that a sender meets the computerized high-type receiver is equal to 1/8 in every repeated game, which is common knowledge in the experiment.

¹⁰ More precisely, the effect on joint payoffs is ambiguous except if μ is sufficiently close to 1, in which case the sender chooses the conditionally trusting strategy that starts with H in both games, and thus, equilibrium behavior and trust outcomes are identical.

¹¹ Many experiments on hidden-information games involve such players (see, e.g., Neral and Ochs, 1992; Andreoni and Miller, 1993; Anderhub et al., 2002; Grosskopf and Sarin, 2010; and Embrey et al., 2015).



Notes: The sender's payoffs that change in the *hr*- and *le*-treatments in comparison to the *bl*-treatments appear in boxes.

Fig. 3. The stage-game payoffs of all experimental games.

Table 1
Experimental design.

Treatment	#high-type computers	#myopic computers	μ	μ_B	ρ_B	# Subjects	# Sessions
<i>bl</i>	1	0	$\frac{1}{8}(7\rho + 1)$	0.158	0.038	90 + 90	6 + 6
<i>hr</i>	1	0	$\frac{1}{8}(7\rho + 1)$	0.256	0.150	90 + 90	6 + 6
<i>le</i>	1	0	$\frac{1}{8}(7\rho + 1)$	0.333	0.238	90 + 90	6 + 6
<i>vcle</i>	1	3	$\frac{1}{8}(4\rho + 1)$	0.333	0.417	72 + 72	6 + 6

Notes: As explained in the text, μ equals $\frac{1}{8} + \frac{7 - C_{myopic}}{8}\rho$, where ρ is the share of experimental subjects that behave like a high-type receiver, and $C_{myopic} \in \{0, 3\}$ refers to the number of computerized receivers that always default; μ_B is the value of μ that makes the sender in the binary version Γ_B indifferent between the non-trusting strategy “always N” and the conditionally trusting strategy that starts with H; and ρ_B is the threshold value of ρ so that $\mu = \mu_B$.

We have four pairs of gradual and binary game treatments, and Fig. 3 presents experimental stage-game payoffs in these treatments. Note that the receiver payoffs presented in Fig. 1 do not change across treatment pairs. The receiver stage-game payoff values and $\delta_R = 0.5$ are such that if a subject in the role of receiver is a monetary payoff maximizer, then this subject is a *low-type receiver* given our definition in Section 3. More specifically, given these experimental values, a receiver who maximizes own monetary payoff will *never* choose R in the binary game and may choose R *only* after L in order to imitate a high type but will default once M or H is chosen. However, we cannot fully control the preferences and choices of subjects, and therefore, $\mu = 1/8$ in the experiment if and only if the behavior of subjects in the receiver role is consistent with low-type behavior. Put differently, if some subjects in the receiver role exhibit high-type behavior, this will increase the experimental magnitude of μ above $1/8$.

As discussed in the previous section and stated in Proposition 1, Γ_G promotes trust building and efficiency relative to Γ_B if μ is relatively low, but may hinder efficiency if $\mu > \mu_B$. We designed four pairs of experimental treatments that manipulate μ and μ_B in order to test predictions of Proposition 1. The experimental variation in μ_B across pairs of treatments stems from the variation in sender payoffs. We vary μ experimentally via different types of computerized receivers. We now describe our experimental treatments.

- *bl-treatments*: These are the *baseline* treatments that were presented earlier in Fig. 1. The computerized high-type receiver is the only computerized player in these treatments.
- *hr-treatments*: These treatments only modify two of the sender payoffs vis-à-vis the baseline treatments making trust more risky (payoff values in boxes in Fig. 3 indicate the modified payoffs): if the sender chooses H in the (binary or gradual) game and the receiver defaults, then the payoff of the sender is -20 (instead of 0 as in the *bl*-treatments). In order to make the sender payoffs more balanced across L, M and H, we also made the following adjustment: if the sender chooses M and the receiver defaults, then the sender payoff is 4 (instead of 12 as in *bl*). We denote these *high-risk* games as the *hr*-gradual game and the *hr*-binary game. These payoff changes increase the theoretical μ_B as indicated in Table 1.

- *le-treatments*: These treatments reduce efficiency gains from mutual cooperation by modifying sender payoffs in the *bl*-treatments as follows. If the sender chooses L (M) [H] and the receiver returns, the sender receives a payoff of 28 (32) [36], rather than 32 (44) [56] as in *bl*. We denote these *low-efficiency* games as the *le*-gradual game and the *le*-binary game. These changes increase the theoretical μ_B further as indicated in Table 1.

- *vcle-treatments*: These treatments are identical to the *le*-treatments except that, in addition to the computerized high-type receiver, three out of eight receivers are computerized and programmed to always default, which is common knowledge. We denote these *low-efficiency* games with *various computerized receivers* as the *vcle*-gradual game and the *vcle*-binary game. Thus, μ_B is identical across *le*-treatments and *vcle*-treatments, however μ is lower in the *vcle*-treatments than in others as indicated in Table 1. Below, we explain this difference in μ in detail.

Difference in μ across treatments: Due to the presence of one computerized high-type receiver in all treatments and the additional presence of three computerized receivers that always default in the *vcle*-treatments, the experimental magnitude of μ is given by the following equation:

$$\mu = \frac{1}{8} + \frac{7 - C_{myopic}}{8} \rho, \tag{1}$$

where ρ is the share of experimental subjects that behave like a high type, and $C_{myopic} \in \{0, 3\}$ refers to the number of computerized receivers that always default.¹² Thus, μ is strictly lower in the *vcle*-treatments than in others provided that $\rho > 0$ and assuming ρ is constant across treatments. By (1) above, whether μ is above or below μ_B in our treatments depends on ρ , which in turn depends on subjects' preferences and behavior. We define ρ_B as the theoretical threshold value of ρ so that $\mu = \mu_B$; that is, ρ_B solves $\frac{1}{8} + \frac{7 - C_{myopic}}{8} \rho_B = \mu_B$. Hence, ρ_B is the theoretical value quantifying the expected benefit of Γ_G over Γ_B in our treatments because

$$\rho > \rho_B \text{ if and only if } \mu > \mu_B.$$

This implies that the higher the magnitude of ρ_B , the higher the chances we have $\rho < \rho_B$, (i.e., $\mu < \mu_B$), and the higher the chances Γ_G generates higher joint payoffs than Γ_B . Conversely, the lower the magnitude of ρ_B , the higher the chances $\rho > \rho_B$, (i.e., $\mu > \mu_B$), and the higher the chances joint payoffs in Γ_B exceed those in Γ_G (see μ , μ_B , and ρ_B values for each pair of treatments in Table 1). Hence, we derive Hypothesis 1 indicating that the comparison of payoffs across the gradual game and its binary version depends on ρ_B .¹³

Hypothesis 1. (i) The gradual game format is more likely to be harmful when ρ_B is low, and more likely to be beneficial when ρ_B is high. More generally, **(ii)** the difference in (sender, receiver, and joint) payoffs between a gradual game and the corresponding binary game is increasing in ρ_B because the predicted prevalence of “*always N*” in the binary game is increasing in ρ_B .

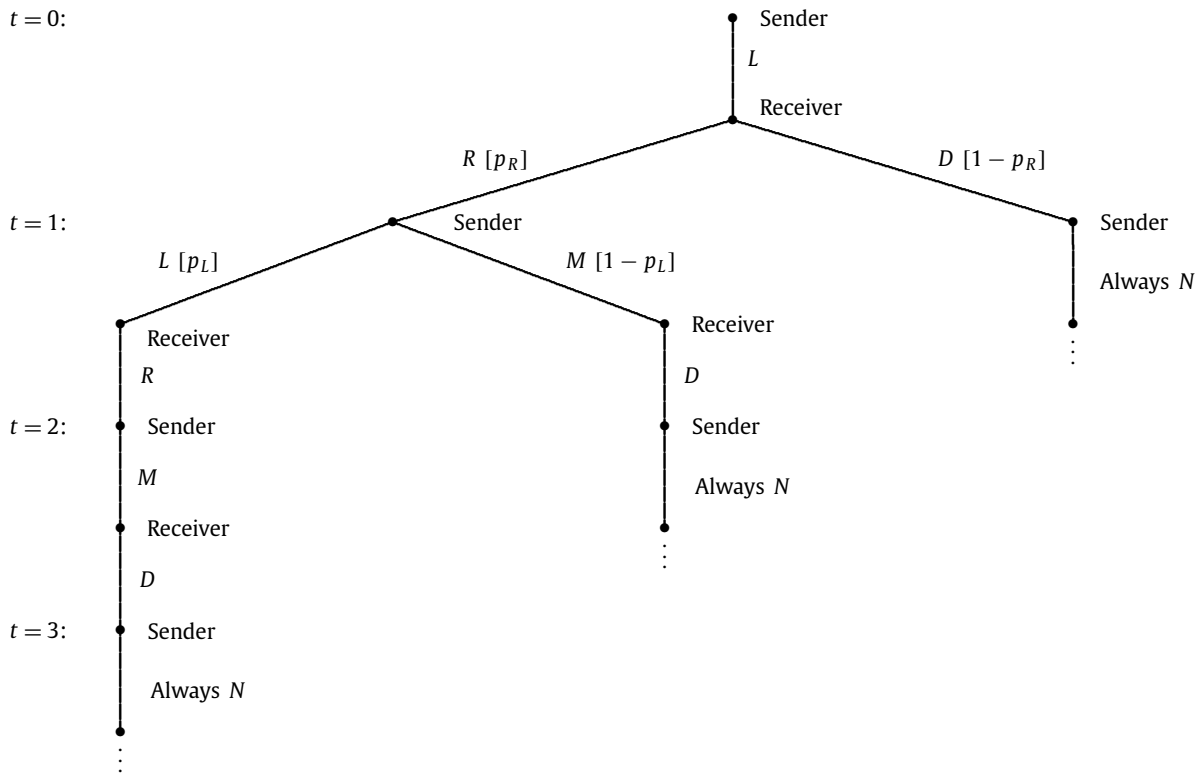
To explain Hypothesis 1 intuitively, consider the benchmark case where $\rho = 0$; that is, subjects in the role of a receiver are all low-type receivers and therefore, $\mu = 1/8$. If $\mu = 1/8$, then trust is too risky for senders in our experimental binary games, and the unique equilibrium is “*always N*”. In contrast, every equilibrium of the experimental gradual games involves trust as discussed below and thus, we are in part (ii) of Proposition 1. But $\mu = \frac{1}{8} + \frac{7 - C_{myopic}}{8} \rho$, and thus, μ exceeds 1/8 in the experiment if $\rho > 0$. So, how robust is the “*always N*” prediction in the binary game? This is summarized by ρ_B : the higher the ρ_B value, the higher the sender’s risk when trusting in the binary game, and the more robust the “*always N*” prediction. More formally, the magnitude of ρ_B equals the minimum fraction of subjects who must behave like a high-type receiver so that the conditionally trusting strategy that starts with H is the unique equilibrium of the binary game. Thus, ρ_B can be interpreted as a statistic reflecting the sender’s risk of trusting in the binary game. As an example, the ρ_B value is only 3.8% in the baseline treatments meaning that if more than just 3.8% of experimental subjects behave like a high type, then trust becomes part of equilibrium in the baseline binary game, and part (ii) of Proposition 1 is not relevant. In

¹² There are two possible, but not mutually exclusive, explanations regarding subjects in the role of a receiver who deviate from own payoff maximization and behave like a high type. Firstly, a fraction of subjects may be prosocial (see Kartal and Müller (2018) regarding the relevance of social preferences in infinitely repeated settings). Secondly, cognitive limitations may play a role (we are grateful to an anonymous referee for this suggestion). That is, subjects may misinterpret payoff parameters and incentives of the game and believe trustworthy behavior maximizes their payoff. In accordance with either explanation, senders may then have so-called “homemade beliefs” (Camerer and Weigelt, 1988) representing the percentage of receivers who are experimentally assigned a low type but choose R rather than D at any trust level. We revisit this issue in Section 6.

¹³ To show how ρ_B can be computed, take as an example the parameters of the *bl*-treatments. We first find the value of μ_B , which solves

$$(1 - \mu_B) \left(0 + 0.75 \left(\frac{24}{1 - 0.75} \right) \right) + \mu_B \left(\frac{56}{1 - 0.75} \right) = 96,$$

i.e., the sender of the binary game is indifferent between “*always N*” and the conditionally trusting strategy which starts with H . Since $\mu_B = 0.158$, and $\mu_B = 1/8 + \rho_B(7/8)$, we find $\rho_B = 0.038$.



Notes: As an example, $p_R = 3/11$ and $p_L = 34/43$ in the *bl*-treatments. See the text for the full description of PBE with both types of receivers.

Fig. 4. The equilibrium path of *E* with a low-type receiver.

this case, gradualist strategies are unlikely to be useful and may backfire. In contrast, if ρ_B is high, such as in the *vcl*-treatments, then trust is very risky for the sender in the binary game, and the gradual game is likely to boost trust and efficiency relative to its binary version. To summarize: the higher the ρ_B value, (1) the higher the sender's risk of trusting and the lower the rate of trust in the binary game, and (2) the higher the relative efficiency of the gradual game, giving rise to Hypothesis 1.

We now discuss the perfect Bayesian Nash equilibria (PBE) of the gradual games with the parametrization mentioned above. Again, we begin by assuming that $\rho = 0$. PBE predictions tend to be much more robust to $\rho > 0$ (i.e., $\mu > 1/8$) in the gradual games than in their binary versions. Still, we will discuss the equilibrium implications of μ values substantially higher than $1/8$ for completeness at the end of this section.

Every PBE of every gradual game in our experiment is gradualist, and all PBEs are qualitatively identical at $t = 0$ and $t = 1$. In every equilibrium, the sender starts the relationship with *L*, weakly increases the trust level as long as the receiver returns, and punishes default with static Nash reversion (i.e., always *N*). In every PBE, *H* is eventually reached after a testing phase if and only if the receiver is high type. A formal and complete description of the PBEs of the experimental gradual games as well as the related proofs can be found in Online Appendix A.1. As all PBEs of every gradual game are similar, below we focus for the sake of simplicity on a specific equilibrium, which we denote by *E*. This equilibrium exists in every gradual game of our experiment. After explaining *E*, we briefly discuss how other PBEs differ. We focus on *E* because it is Pareto efficient and provides the fastest information revelation. In particular, the sender and the receiver randomize only once on the equilibrium path of *E* unlike in other PBEs. As the high-type receiver always returns, we describe only the equilibrium path of *E* with a low-type receiver below (see also Fig. 4).

- (i) At $t = 0$, the sender chooses low trust (*L*), and the receiver randomizes between *R* and *D*.
- (ii) If the receiver chooses *D* at $t = 0, 1, \dots$, the sender strategy at $t + 1$ is “always *N*.”
- (iii) At $t = 1$, the sender randomizes between *L* and *M* if the receiver chooses *R* at $t = 0$.
 1. If the outcome of the sender randomization at $t = 1$ is *L*, the receiver chooses *R* with certainty. At $t = 2$ the sender chooses *M*, and the receiver chooses *D* with certainty. Play continues according to (ii).
 2. If the outcome of the sender randomization at $t = 1$ is *M*, the receiver chooses *D* with certainty. Play continues according to (ii).

Given (i)–(iii) above, the sender learns the type of the receiver with certainty at the latest by the end of $t = 2$. In equilibrium E as well as in every other PBE, the choice of M always precedes H , and the low type optimally defaults whenever M is chosen.¹⁴ Thus, in every PBE the sender eventually chooses H if and only if the receiver is high type. While the low-type receiver immediately defaults once M is chosen, he always returns with strictly positive probability if L is chosen on the equilibrium path. This is true in every PBE, and the only distinction across various equilibria of the gradual game is numerical at $t = 0$ and $t = 1$ as mentioned before. That is, in the notation of Fig. 4, only p_R and p_L can vary across equilibria at $t = 0$ and $t = 1$ respectively. At $t = 2$, E prescribes M with certainty if L and R were chosen at both $t = 0$ and $t = 1$ as seen in Fig. 4, whereas other PBEs prescribe randomization between L and M . As a result, full information revelation can take longer in PBEs other than E . All of these statements apply to every gradual game of the experiment. Our second hypothesis concerns the equilibrium predictions for the sender behavior in the gradual game treatments.

Hypothesis 2 (Gradualism). Sender behavior is consistent with the testing strategies described by the PBEs of the gradual games.

The equilibrium predictions we described above are robust to values of μ greater than $1/8$ up to a certain threshold specific to each gradual game. If μ exceeds the respective threshold in a game, the best response of the sender is to choose M at $t = 0$ rather than L , in contrast to our PBE description above. In addition, once μ is sufficiently high and close to 1, gradualist strategies are no longer used in equilibrium as it is optimal for the sender to choose H at $t = 0$.

4.1. Experimental protocol

Our design is between-subject: each subject participated in only one treatment. In all treatments, subjects participated anonymously in a sequence of indefinitely repeated games. We refer to each indefinitely repeated game as a “match” and each repetition of the stage game within a match as a “round.” At the beginning of a match, each sender was randomly matched with a receiver. In each session, subjects were informed about the number of computerized receivers, and how they are programmed to choose. Senders and receivers were randomly rematched after the end of each match. Each session ended either after 25 matches were completed, or at the end of the first match that was completed after 75 minutes passed, whichever occurred first.

We implemented an indefinitely repeated game in the lab by using a random continuation rule. The probability of continuation after each round was the same for all matches in all treatments and was equal to the discount factor of the sender $\delta_S = 0.75$. Thus, the expected number of rounds in every match is four.

Each subject assigned the receiver role had a discount factor of 0.5; i.e., $\delta_R = 0.5$. This was implemented as follows. Starting from the second round of a match, the receiver payoffs in Fig. 3 were reduced by $1/3$ in every round. For example, if the match continues to the second round and the sender chooses no trust in the second round, then the receiver obtains 10 in the second round, rather than 15 (since the continuation probability of 0.75 times $2/3$ equals 0.5, the discount factor of the receiver is effectively equal to 0.5). This was clearly explained in the experimental instructions. Hence, our design for implementing $\delta_R = 0.5$ combines the case where the future is less valuable than the present and the case where each interaction randomly terminates with a positive probability. This design is a variation on an already existing experimental method in which pure payoff discounting is followed by pure random termination.¹⁵

The experiment was conducted at the University of Vienna and the University of Konstanz.¹⁶ Subjects were recruited from the student populations. We conducted six sessions for each treatment (48 sessions in total). 684 students participated in the experiment. All sessions were conducted through computer terminals using a program written in z-Tree (Fischbacher, 2007). The average payoff per subject was about €33, which includes €2 for filling in a post-experimental questionnaire. The instructions for the baseline treatments can be found in Online Appendix E.

5. Results

We now present our experimental results. In Section 5.1, we test Hypothesis 1, comparing payoffs between gradual games and their binary counterparts. Consistent with our predictions, we find that the gradual game reduces joint efficiency relative to its binary version in the treatment pair with the lowest risk of trusting in the binary game (i.e., lowest ρ_B), whereas the gradual game increases joint efficiency in the treatment pair with the highest risk (i.e., highest ρ_B). In the treatment pairs with intermediate levels of ρ_B , the difference in joint payoffs across the two types of games is not statistically significant.

To check that our results in Section 5.1 are driven by the mechanisms in our theory, in Section 5.2 we study individual behavior. More specifically, we study the extent of gradualist behavior that is consistent with equilibrium (Hypothesis 2), and

¹⁴ We opted for a design in which receiver and sender payoffs vary in a *nonlinear* fashion because it ensures the existence of a relatively simple equilibrium such as E in which defaulting after M is the dominant strategy for the receiver, and both players randomize only once on the equilibrium path.

¹⁵ To be more precise, this existing method involves a fixed, known number of rounds played with *certainty*, and the payoffs of these rounds are discounted at a known discount rate of $\delta < 1$. After the rounds with certainty are played, payoffs are no longer discounted and in every round, the interaction continues to the next round with probability δ . See Fréchet and Yuksel (2017), and the references therein.

¹⁶ In particular, corresponding pairs of gradual and binary game experiments were always conducted at the same university.

Table 2
Summary statistics of efficiency indices by treatment.

Treatment	Joint-efficiency index	Sender-efficiency index	Receiver-efficiency index	ρ_B
<i>bl</i> -gradual game	0.332 (0.020) <**,***	0.079 (0.019) ≈	0.774 (0.038) <**,***	
<i>bl</i> -binary game	0.456 (0.031)	0.121 (0.050)	1.055 (0.089)	0.038
<i>hr</i> -gradual game	0.266 (0.021) ≈	0.039 (0.036) ≈	0.668 (0.039) <**,***	
<i>hr</i> -binary game	0.314 (0.040)	0.032 (0.057)	0.898 (0.085)	0.150
<i>le</i> -gradual game	0.380 (0.015) ≈	0.046 (0.029) ≈	0.607 (0.026) ≈	
<i>le</i> -binary game	0.384 (0.058)	0.033 (0.058)	0.648 (0.079)	0.238
<i>vcle</i> -gradual game	0.130 (0.014) >***,***	−0.011 (0.013) >*,**	0.454 (0.039) ≈	
<i>vcle</i> -binary game	0.012 (0.019)	−0.135 (0.048)	0.412 (0.075)	0.417

Notes: This table reports average efficiency indices $(\pi^{obs} - \pi^N)/(\pi^C - \pi^N)$ as defined in the text. Standard errors clustered by session are in parentheses. Test results are shown in the form $\chi^{a,b}$, where $\chi = "<"$ indicates that the summary statistic in the line above is smaller than the summary statistic in the line below, $\chi = "≈"$ indicates that there is no significant difference at the 10%-level, a represents the significance level according to a Mann-Whitney U test, and b denotes the significance level according to the regression-based analysis in Online Appendix D. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

check that sender trust in the binary game varies with ρ_B as predicted. We find that gradualism is a robust phenomenon as many senders employ the gradualist strategies consistent with equilibrium in all four gradual games. In addition, we find that sender behavior strongly responds to the variation in the risk of trusting in the binary games. Thus, while trust is prevalent in all four gradual games, sender trust varies starkly across their binary counterparts giving rise to the efficiency differences reported in Section 5.1.

In our analysis, we focus on experienced play and discard the data of the first ten matches.¹⁷ We use a combination of nonparametric and parametric methods in our tests. For our nonparametric tests, the unit of observation is the session average, and for all tests we report two-sided significance levels.

5.1. Payoff comparisons across games and treatments

We now explain the analysis and results in detail. Since the length of each match is random in the experiment, and the match length affects payoffs, we control for the number of rounds played in each match in our payoff comparisons. In particular, we calculate the efficiency index

$$\frac{\pi^{obs} - \pi^N(r)}{\pi^C(r) - \pi^N(r)},$$

where π^{obs} denotes the realized payoff of a match in a treatment, r denotes the number of rounds in that match, $\pi^N(r)$ denotes the payoff if the static Nash equilibrium is played in all r rounds, and $\pi^C(r)$ denotes the payoff from repeated full cooperation (i.e., in every round, the sender chooses H , and the receiver chooses R).

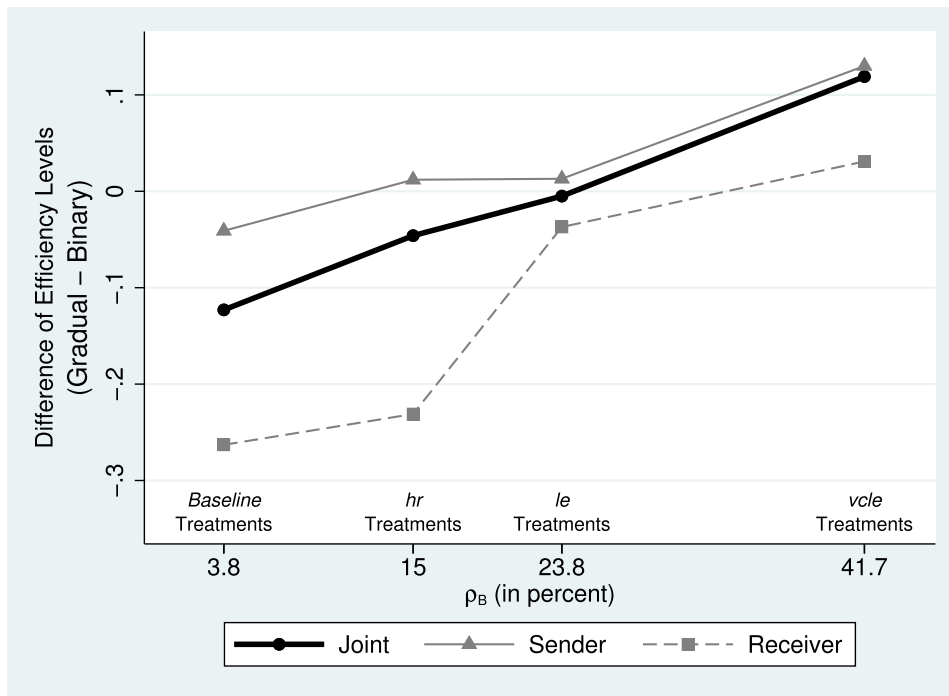
We compute three efficiency indices: one for joint payoffs, one for sender payoffs, and the third one for receiver payoffs.¹⁸ These indices are meant to measure the cooperativeness of play within a match. In particular, all three indices equal one if players coordinate on the most cooperative outcome, and zero if static Nash play prevails.¹⁹ Table 2 reports the summary statistics of the three indices by treatment pair and type of game (binary or gradual) along with statistical test results. The joint-efficiency index is clearly above zero for all treatments except for the *vcle*-binary game. This is because sender trust is limited in the *vcle*-binary game (consistent with the high ρ_B parameter relative to other treatment pairs) as we will discuss in the next section. A further look at the respective indices for senders and receivers reveals that in every pair of treatments, the sender-efficiency index is considerably lower than the corresponding receiver-efficiency index, which suggests that many receivers gain at the expense of senders by (eventually) defaulting.

Table 2 shows that in the *bl*-treatments, the joint-efficiency index is 12.3 percentage points higher, and the receiver-efficiency index is 26.3 percentage points higher in the binary game than in the gradual game, which is statistically

¹⁷ The percentage of behavior consistent with equilibrium predictions increases considerably during the first ten matches. Online Appendix B.1 provides an analysis of behavior in the first ten matches.

¹⁸ Receiver-efficiency indices are calculated only for subjects in the role of a receiver. When a sender is matched to a computerized receiver, the joint-efficiency index is computed exclusively on the basis of the sender payoffs.

¹⁹ Note that the receiver-efficiency index may exceed one, since the efficiency gains from high trust is shared very unequally if the receiver defaults. In particular, the receiver-efficiency index for a receiver that defaults after the sender chooses H in the first round is above 1 in matches that last 1-3 rounds. As an example, it equals 2.4 in a match that lasts one round.



Notes: This Figure plots the difference in efficiency levels between a pair of gradual and binary games against the corresponding level of ρ_B . “Joint,” “Sender,” and “Receiver” refer to the joint-efficiency index, the sender-efficiency index and the receiver-efficiency index, respectively.

Fig. 5. The relationship predicted by Hypothesis 1 part (ii).

significant (in both cases, $p < 0.05$ according to a Mann-Whitney test).²⁰ As discussed in Section 4 (see also Table 2), this is the treatment pair associated with the lowest value of ρ_B . Recall the interpretation of $\rho_B = 0.038$: the unique equilibrium prediction of the *bl*-binary game is “always N” if $\rho < \rho_B = 0.038$, where ρ is the share of experimental subjects who behave like a high type receiver (against their monetary interest). Once more than 3.8% of subjects in the receiver role behave like a high type, trust is sustained in the equilibrium of the *bl*-binary game and can be expected to be prevalent in practice. This is indeed the case: the data suggests that ρ is higher than 3.8%, as sender trust and receiver return are prevalent in the *bl*-binary game and hence, the gradual game format is associated with reduced efficiency in the *bl*-treatments. As the respective ρ_B value increases to 0.15 and 0.238 in *hr* and *le*, the difference in the joint-efficiency index between the gradual and binary games becomes statistically insignificant in both treatments. Still, the receiver-efficiency index is lower in the *hr*-gradual game than in the *hr*-binary game ($p < 0.1$ according to a Mann-Whitney test). Finally, in *vcl* where ρ_B is at its highest, gradualism is beneficial: consistent with the high theoretical value of ρ_B , trust is infrequent in the *vcl*-binary game, whereas the use of gradualist strategies in the *vcl*-gradual game boosts efficiency (e.g., $p < 0.01$ for joint payoffs according to a Mann-Whitney test). Thus, our findings are in line with part (i) of Hypothesis 1: the performance of the gradual game relative to its binary version is the worst in the treatment in which ρ_B is lowest (*bl*), where it is harmful to efficiency. In contrast, the gradual game significantly improves efficiency in the treatment where ρ_B is highest (*vcl*).

More generally, the difference in payoffs between a gradual game and its binary version (i.e., the relative efficiency of the gradual game) is increasing in ρ_B as predicted in part (ii) of Hypothesis 1. This is shown in Fig. 5. We now test that the increasing pattern in Fig. 5 has statistical support by regressing the efficiency indices on a dummy for the gradual game (“gradual”) indicating whether an efficiency-index value comes from a gradual game or a binary game, the ρ_B value for the respective treatment, and interaction terms for the gradual game and the respective ρ_B value (“gradual \times ρ_B ”). Doing so, we provide evidence for part (ii) of Hypothesis 1 as follows (see Table 3):

- The coefficient on ρ_B is negative and significant ($p < 0.01$ for all three indices) indicating that there is a significant decline in efficiency in the binary game as ρ_B increases consistent with our theory.
- The coefficient on “gradual \times ρ_B ” is positive indicating that as predicted the relative efficiency of the gradual game is increasing with ρ_B for all three efficiency indices ($p < 0.01$ for joint and receiver payoffs, and $p < 0.05$ for sender payoffs).

To summarize, we find compelling evidence that the impact of the gradual game relative to its binary version is consistent with our theoretical framework and Hypothesis 1. Some other remarks are in order. Firstly, our efficiency comparisons

²⁰ Table 2 indicates that the difference is also statistically significant according to our parametric approach explained in more detail in Online Appendix D ($p < 0.01$).

Table 3
Testing Hypothesis 1 part (ii).

	Joint-efficiency index	Sender-efficiency index	Receiver-efficiency index
gradual	−0.146*** (0.0392)	−0.0643 (0.0477)	−0.323*** (0.0808)
ρ_B	−1.100*** (0.110)	−0.649*** (0.165)	−1.780*** (0.287)
gradual \times ρ_B	0.622*** (0.140)	0.426** (0.175)	0.945*** (0.314)
Constant	0.524*** (0.0319)	0.150*** (0.0436)	1.124*** (0.0740)
Observations	5,592	5,592	4,365
R^2	0.094	0.018	0.097

Notes: Robust standard errors clustered by session in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

may seemingly suggest that equilibrium gradualist strategies are sometimes futile or even harmful *within* the gradual game itself; however, this is not the case. As will be discussed in Section 5.2.1, equilibrium behavior generates the highest expected payoff for the senders of the gradual game given the empirical receiver behavior. Secondly, the gradual game improves sender payoffs relative to its binary version only in the *vcle*-treatments, whereas the improvement could theoretically be expected in the other treatment pairs as well. However, we observe that receivers tend to be more trustworthy in the binary game than in the gradual game, which may explain why sender payoffs differ across the gradual and binary games only when the asymmetric information problem is most severe as in the *vcle* treatments. We elaborate on this finding in Section 6.

5.2. Actions and strategies

Below, we first describe subject behavior in the gradual games paying particular attention to the degree to which senders use equilibrium strategies in the gradual games to test Hypothesis 2. Next, we study sender behavior in the binary games and how it varies with ρ_B . The link between our efficiency results and the observed sender behavior is discussed in detail in Section 5.2.2.

5.2.1. Behavior in the gradual game

The primary purpose of this section is to test whether or not senders play the gradualist strategies as predicted by Hypothesis 2. We test this hypothesis in two ways.

- We test if equilibrium predictions outlined in Section 4 *predict outcomes better than alternative strategies*. When looking only at the first round of a match, every equilibrium predicts *L*, so we simply test whether this is chosen more frequently than the other three choices, namely *N*, *M*, and *H*. When considering later rounds in addition, we define a number of strategies, which are both intuitively reasonable and feature heavily in our data, against which we compare our equilibrium predictions.
- We test if our theory has *predictive power*, i.e., if senders are more likely than chance to play an equilibrium strategy. We also test the performance of other strategies we identified in the data against chance.

Throughout this section, all statistical comparisons are made using two-sided *binomial tests*, treating each session as an independent observation: an observation counts as a “success” if equilibrium strategies are more frequent in that session than predicted by chance (when testing predictive power), or if equilibrium predicts better than the alternative against which it is being compared (when testing equilibrium against other strategies).

We conduct tests separately for each treatment as well as pooling data across all four treatments. In pooling the data, we are testing the predictive success of our theory in a class of games, rather than a single game. This is, in a sense, a more robust test of the theory. It also gives us far more statistical power to make comparisons between different types of sender strategies.²¹

Behavior in the first round of a match Table 4 reports subject behavior in the first round of a match. The equilibrium choice of low trust (*L*) is by far the most common choice in the first round of all gradual games, with 53.5 – 68.5% of senders choosing *L* in the first round of a match, depending on the treatment. The minimum percentage of *L* in any session is 40%, so *L* is chosen with a significantly greater frequency than chance in the pooled data ($p < 0.01$) and in all four treatments individually ($p = 0.031$ in each treatment). Importantly, no other trust level is chosen with a significantly greater frequency

²¹ When we look separately at each treatment, a two-sided binomial test result is statistically significant only if either the hypothesis in question holds in all six sessions (in that case, $p = 0.031$ in a two-tailed test) or the hypothesis holds in five sessions with a tie in one session (in that case, $p = 0.062$ in a two-tailed test).

Table 4
Sender and receiver behavior in the first round of a match in the gradual games.

Treatment	Sender choice			
	No trust (N)	Low trust (L)	Medium trust (M)	High trust (H)
<i>bl</i>	7.3%	67.1%	16.5%	9.1%
<i>hr</i>	19.9%	58.5%	13.7%	7.9%
<i>le</i>	19.7%	53.5%	11.5%	15.3%
<i>vcl</i>	27.4%	68.5%	3.5%	0.7%
Receiver return rates (including computers)				
<i>bl</i>	–	64.2%	39.1%	33.3%
<i>hr</i>	–	58.3%	34.8%	35.3%
<i>le</i>	–	71.2%	55.4%	58.2%
<i>vcl</i>	–	41%	48%	20%
Receiver return rates (excluding computers)				
<i>bl</i>	–	59.2%	28.6%	23.6%
<i>hr</i>	–	52%	28.4%	23.3%
<i>le</i>	–	66.9%	47.9%	51.6%
<i>vcl</i>	–	58.1%	83.3%	0%

than chance. In addition, *L* is the most common first-round trust choice in 23 out of 24 sessions ($p < 0.01$ in the pooled data, and $p = 0.031$ in every treatment but *vcl*). Table 4 also shows that receivers' first-round return rate is at its highest if the sender chooses *L* in every treatment except for *vcl*, but this is noisy as only 1.7% of subjects in the role of a receiver in *vcl* observe a choice of *M* in the first round. The relatively high rate of returning low trust is consistent with equilibrium. As discussed in Section 4, a low-type receiver has an incentive to return after *L* to build up a reputation and default once the trust level increases.²²

Behavior analysis involving later rounds of a match To continue testing Hypothesis 2, we now classify senders' behavior based on the outcomes in the first two, three, and four rounds of a match. In order to provide a classification of sender behavior in the first $n \geq 2$ rounds, we must restrict the data to matches that last at least n rounds. In view of this requirement, we limit our analysis to the first four rounds. Classifying sender behavior in the first five rounds, for example, requires restricting the data to matches that last at least five rounds, which results in the loss of a large majority of the data and thus small numbers of observations due to the random continuation rule we use to implement indefinitely repeated games in the lab. We begin our analysis by identifying in the data "gradualist" game histories defined as follows.

Gradualist: Sender behavior is gradualist if the sender starts the relationship with either *L* or *M*, (weakly) increases the trust level as long as the receiver returns, and punishes default with no trust thereafter.

We further classify "gradualist" behavior according to whether or not it is consistent with our equilibrium predictions (denoted by "equilibrium" and "non-equilibrium"). The proportion of "equilibrium" behavior is our main variable of interest. Of course, we do not observe the entire sender strategy in the experiment, and therefore, we can only rely on observed histories of sender behavior for equilibrium classification. Since every equilibrium strategy predicts randomizing between two actions in some rounds, we say that the sender behavior satisfies the criterion of being *equilibrium* if either one of these actions is chosen.²³ In addition, to be classified as *equilibrium*, sender behavior must always be a best response to actual receiver behavior (which may be on or off the equilibrium path).

Among game histories that are not consistent with our (*equilibrium* or *non-equilibrium*) gradualist definition, many can be classified as being consistent with four intuitive strategies explained below.

Always N: The sender always chooses *N* in which case the receiver has no choice to make.

H until default: The sender starts by choosing *H* in the first round and chooses *H* if the receiver returned in all previous rounds, and *N* otherwise.

Lenient: The sender behavior is lenient if the sender starts with a positive trust level, (weakly) increases the trust level as long as the receiver returns, and punishes default leniently by either choosing a strictly lower but positive trust level or choosing *N* for a limited number of rounds followed by a positive trust level.

Hybrid: The sender behavior is hybrid if the sender starts with "always *N*" and then switches to "equilibrium" in the second, third or fourth round. This behavior suggests that the sender is roughly indifferent between the two strategies at the beginning of the game. An alternative name for this behavior would be suspicious gradualist, akin to suspicious tit-for-tat (Boyd and Lorberbaum, 1987).

²² We administered a questionnaire after the experiment and asked senders and receivers about their incentives to trust and return, respectively. Consistent with the theory, many receivers stated that if a low trust level was chosen, they imitated the computerized player with the hope that the sender would choose a higher trust rate in the coming rounds.

²³ For example, focusing on match histories up to (and including) the sender's choice in the second round, three are consistent with equilibrium: in all three, the sender begins with *L* at $t = 0$, then chooses either *L* or *M* at $t = 1$ if the receiver chose *R* at $t = 0$, and *N* if the receiver defaulted at $t = 0$.

Table 5
Classification of sender behavior in the gradual games in the first n rounds, $n = 2, 3, 4$.

Treatment	Matches that last at least n rounds	Gradualist		Other				
		Equilibrium	Non-equilibrium	Always no trust	High trust until default	Lenient	Hybrid	Un-classified
<i>bl</i> -gradual	$n = 2$	55.8%	19.4%	2.1%	5.4%	3.7%	5.6%	8.1%
	$n = 3$	47.7%	18.8%	2.3%	5.2%	9.1%	3.1%	13.8%
	$n = 4$	38.3%	17.4%	1.9%	4.9%	16.3%	4.5%	16.7%
<i>hr</i> -gradual	$n = 2$	46.3%	9.1%	13.4%	6.3%	4.7%	8.4%	11.9%
	$n = 3$	41.3%	9.6%	9.0%	5.8%	8.1%	7.8%	18.3%
	$n = 4$	37.1%	11.4%	8.5%	5.5%	9.9%	4.4%	23.2%
<i>le</i> -gradual	$n = 2$	48.3%	9.2%	15.1%	13.7%	1.4%	4.7%	7.6%
	$n = 3$	36.7%	16.1%	12.8%	10.3%	4.2%	6.7%	13.3%
	$n = 4$	30.1%	18.4%	11.0%	9.6%	6.3%	7.7%	16.9%
<i>vcle</i> -gradual	$n = 2$	57.7%	8.8%	21.9%	0.8%	0.4%	5.4%	5.0%
	$n = 3$	52.9%	12.2%	17.0%	0.8%	4.0%	5.6%	7.4%
	$n = 4$	50.4%	11.2%	15.2%	0.9%	7.6%	4.9%	9.8%

Note: This table reports the results of the classification of sender behavior in the gradual game treatments as defined in the text. To classify individual sender behavior in the first $n \in \{2, 3, 4\}$ rounds of a match requires discarding data from matches that last shorter than n rounds.

Table 6
Chances of histories generated by random play.

	$n = 2$	$n = 3$	$n = 4$
Equilibrium	10.9	3.3	0.8
Non-equilibrium Gradualist	14.8	5.6	1.7
Always No Trust	3.9	0.6	0.1
High Trust until Default	7.0	1.5	0.3
Lenient	9.4	12.9	10.1
Hybrid	3.9	2.3	0.9

Notes: This table shows chances (in percentage terms) that random play generates match histories consistent with a listed strategy for $n = 2, 3, 4$.

We label sender behavior that does not fit into any one of these categories as “unclassified.” Our classification results are shown in Table 5.²⁴ A first look at this table shows that equilibrium play is common, and by far the most frequent category in all treatments (46.3-57.7% for $n = 2$, 36.7-52.9% for $n = 3$, and 30.1-50.4% for $n = 4$).²⁵ No other strategy comes close for any considered n .²⁶

We begin by testing the performance of equilibrium as well as the other strategies we identified in the data against chance. The first column in Table 6 presents the strategies explained above, and the remaining columns indicate the theoretical chance that random play generates a match history consistent with a given strategy for all considered n . To test the performance of a strategy against chance for given n , the empirical frequency of play consistent with that strategy is tested against the respective value in Table 6.²⁷ Note that the numbers presented in Table 6 are also associated with the “precision” of a theory. In particular, lower values are associated with higher precision because a theory that is consistent with all possible histories of play would account for all the data but is clearly of no use in making predictions. The low figures for equilibrium play for all considered n imply that the precision of equilibrium predictions is high, especially in analysis involving longer rounds.

We find that equilibrium play is more common than chance in *all* sessions for *all* considered n , meaning that our theory has predictive power overall ($p < 0.01$) and in each individual treatment ($p = 0.031$). Importantly, the same is *not* true for any other strategy. The only other strategies that pass the test of chance in *some* treatments (or in the pooled data) for *some* $n \in \{2, 3, 4\}$ are (i) non-equilibrium gradualist in *le* for $n = 3$ and *bl* for $n = 4$ ($p = 0.031$ in each case) as well as the pooled

²⁴ We concentrate on the frequency of equilibrium behavior identified in Section 4 throughout, but this emphasis does not mean that any other behavior observed in the data cannot be rationalized. By invoking heterogeneous sender priors or risk aversion, it is possible to rationalize certain sender strategies such as “*H* until default”, “always *N*” or the non-equilibrium gradualist strategy that begins with *M* at $t = 0$.

²⁵ Online Appendix B.2 contains complete lists of the frequencies of the observed match histories up to the second, third, and fourth sender decisions as well as the assignment of these match histories to the strategies described above for the baseline treatments as an example. The same analysis for the other treatments is available upon request.

²⁶ Note that the frequency of equilibrium play does not differ statistically across any two gradual games except that the *vcle*-gradual game generates the highest frequency of equilibrium play for $n = 4$ (this is significant at the 5 percent level according to Mann-Whitney tests).

²⁷ More precisely, the numbers in Table 6 are the chance terms we computed taking into account the presence of computerized high-type receivers and are valid for *bl*, *hr*, and *le*. The numbers against which we compare *vcle* data are slightly different as they also account for the presence of computerized receivers programmed to always default.

data for $n = 3$ and 4 ($p < 0.01$); and (ii) always N in le for $n = 2, 3$, and 4 ($p = 0.031$), and the pooled data for $n = 2$ and 3 ($p = 0.064$ and 0.023 respectively).

We now evaluate the success of equilibrium play against other strategies. Pooling data across treatments, we find that equilibrium play is significantly more frequent than each of the other strategies for all considered n ($p < 0.01$ in each case). For example, non-equilibrium gradualist behavior is consistent with many more histories (i.e., less precise) than equilibrium as indicated in Table 6, but still much less prevalent than equilibrium. Comparing equilibrium play against other strategies separately in each treatment, we have 60 pairwise comparisons: five strategies (other than equilibrium) in four treatments for three values of n . We find that equilibrium play is significantly more frequent than other strategies in 55 out of 60 pairwise comparisons ($p = 0.031$ for 52 comparisons and $p = 0.062$ for 3 comparisons). In the rare occasions where the difference in frequency between equilibrium and another strategy is not statistically significant in a treatment, it is always due to failure in *only* a single session, i.e., equilibrium play is more frequent in five out of six sessions even in those exceptions where $p > 0.1$. The exceptions are: non-equilibrium gradualist in bl for $n = 3$ and in le for $n = 4$; always N in vcl for $n = 2$ and 3; and lenient in le for $n = 4$.

In summary, we find strong statistical evidence that our equilibrium predictions have predictive power: equilibrium play is more common than chance in *all* treatments separately as well as in the pooled data for *all* considered n , which does not hold for any other strategy. In addition, equilibrium predictions perform significantly better than alternatives: equilibrium play is statistically the most frequent strategy in the pooled data for all considered n and, looking separately in each treatment, the most frequent in 55 out of 60 pairwise tests. These findings provide support for Hypothesis 2. In addition, sender experience increases the frequency of equilibrium play: the proportion of match histories consistent with equilibrium is larger after the first 10 rounds in all treatments and for all n considered (see Table 8 in Online Appendix B).

We close this section with a discussion of the average realized payoffs of the most commonly played sender strategies in the gradual games, namely the various gradualist strategies, “ H until default,” and “always N .” To be more specific, we compute the average expected return to these strategies (up to the third round) given observed receiver behavior in the experimental sessions.²⁸ Overall, equilibrium behavior provides the highest expected return followed by non-equilibrium gradualist behavior. The worst strategy is “ H until default.” We present a detailed description of the payoff calculations and payoffs of various strategies in Online Appendix C.

5.2.2. Behavior in the binary game

We classify sender behavior in the binary game defining strategies similar to those in the gradual game.²⁹ In particular, our classification involves the following strategies: “ H until default”, “always N ”, “always H ”, “lenient”, and “hybrid”. “Always H ” is similar to the lenient strategy but even more forgiving as the sender always chooses H even if the receiver defaults. The hybrid strategy is such that the sender starts with “always N ” and then switches to “ H until default” in the second, third or fourth round.

Table 7 presents the classification of sender behavior in each binary game. One main feature of the binary game data is the prevalence of trusting behavior except in vcl where the sender’s risk of trusting is the highest. In particular, a majority of senders choose either “ H until default” or “always H ” for all considered n in bl and hr consistent with their relatively low ρ_B values. Furthermore, the use of the strategy “ H until default” is well justified in bl and hr as returning trust is prevalent in both binary games (see Table 17 in Online Appendix C). In contrast, the prevailing preference of senders for gradualist strategies in the gradual games of bl and hr limits efficiency gains relative to their binary versions. As a result, the binary game increases joint payoffs in bl and receiver payoffs in bl and hr as reported in Section 5.1.

The other main feature of the data is that “always N ” is by far the most common strategy only in vcl , which involves the highest ρ_B value as three computerized receivers are programmed to always default. The prevalence of “always N ” results in low payoffs in the vcl -binary game, whereas the widespread use of gradualist strategies in the vcl -gradual game fosters trust and improves efficiency giving rise to findings discussed in Section 5.1.

While we found that the expected sender payoff from the strategy “ H until default” is low in all the gradual games, this is not true in the binary games, with the exception of vcl . This finding relates to the following observation. The receiver is in an identical situation in the first round if the sender chooses high trust in the gradual game or in its binary version. However, we observe that the return rate of subjects in that situation tends to be higher in the binary game than in the gradual game (see Table 15 in Online Appendix B.4).³⁰ In Proposition 1, we assumed that there is no such difference, but if the binary game has a higher share of high-type receivers than the gradual game, this only strengthens part (iv) in Proposition 1 regarding the cost of gradualism.

²⁸ We limit this analysis to three rounds of the more frequently observed strategies to ensure return rates are estimated with a reasonable number of observations.

²⁹ Online Appendix B.3 contains complete lists of the frequencies of the observed match histories (up to the second, third, and fourth round) and their assignment to the strategies described in the text for the baseline treatments as an example. The same analysis for the other treatments is available upon request.

³⁰ Note that the difference in returning behavior across the gradual and binary games develops over time and could therefore stem from the diverging experiences of receivers in the two types of games: testing strategies are not available in the binary game, and senders frequently choose to trust the receiver in three out of four binary games. See also Footnote 31.

Table 7
Classification of sender behavior in the binary games in the first n rounds, $n = 2, 3, 4$.

Treatment	Matches that last at least n rounds	Always no trust	High trust until default	Always high trust	Lenient	Hybrid	Unclassified
<i>bl</i> -binary	$n = 2$	21.6%	52.2%	13.1%	–	12.9%	0.2%
	$n = 3$	16.0%	45.7%	13.6%	3.5%	13.8%	7.4%
	$n = 4$	12.1%	43.0%	12.5%	4.0%	15.1%	13.2%
<i>hr</i> -binary	$n = 2$	24.5%	60.0%	4.6%	–	7.6%	3.3%
	$n = 3$	17.9%	56.5%	5.8%	2.0%	10.5%	7.4%
	$n = 4$	17.3%	53.3%	8.6%	2.7%	9.8%	8.3%
<i>le</i> -binary	$n = 2$	43.0%	44.5%	0.8%	–	10.6%	1.1%
	$n = 3$	38.2%	40.7%	1.4%	0.0%	16.4%	3.4%
	$n = 4$	35.0%	36.1%	0.8%	0.8%	20.8%	6.4%
<i>vcle</i> -binary	$n = 2$	61.0%	25.2%	0.7%	–	12.7%	0.4%
	$n = 3$	52.2%	24.8%	0.0%	1.1%	20.2%	1.8%
	$n = 4$	48.2%	22.3%	0.9%	3.0%	21.6%	4.0%

Note: This table reports the results of the classification of sender behavior in the “Binary” treatments as defined in the text. To classify individual sender behavior in the first $n \in \{2, 3, 4\}$ rounds of a match requires discarding data from matches that last shorter than n rounds.

6. Discussion and concluding remarks

We study strategic behavior in a trust-game setting with an infinite horizon where senders are uncertain about the trustworthiness of receivers. We observe that senders in the gradual games predominantly use equilibrium gradualist strategies in order to identify the “type” of the receiver they are matched with, and in turn, many low-type receivers imitate the high type at low trust levels as predicted by the theory. In other words, the equilibrium form of gradualism is a robust phenomenon in all four gradual games despite the sophisticated structure of equilibrium strategies. Moreover, equilibrium strategies bring senders higher expected payoffs than non-equilibrium strategies.

We compare the performance of the gradual game to its binary version in four pairs of treatments. Among these, the gradual game performs better than its “all-or-nothing” version *only* in the treatment pair where the asymmetric information problem is the most severe (due to the presence of computerized receivers that are programmed to default), and the return to trusting behavior is relatively low for the sender. Otherwise, the binary games generate high levels of cooperation, and gradualism is either unnecessary or even undermines efficiency. The existing literature has emphasized the benefits of gradualism. However, our framework also highlights its possible costs in a broader context, which allows for a comparison to a simple setting that involves dynamic incentives but not the option to use testing strategies.

According to our theoretical model, the comparison of the gradual game to its binary version depends on the share of unconditionally trustworthy, or “high type”, receivers. While rational and materially self-interested subjects in the role of a receiver should act as a low type in our experiment, we observe that a nontrivial share of subjects exhibits high-type behavior by returning trust in the binary game. There are two conceivable reasons for a receiver to exhibit high-type behavior. The first reason is that some receivers are sufficiently prosocial and would prefer returning trust rather than betraying the sender. This explanation is generally supported by a sizeable literature on prosocial preferences as well as the answers subjects gave in our post-experimental questionnaire. Many receivers referred to prosocial motivations and willingness to cooperate as the reason for returning sender’s trust. The second reason is that some subjects who are materially self-interested are also “boundedly rational” and mistakenly believe that they are better off returning trust instead of defaulting especially in the binary game. The two explanations (i.e., prosociality and cognitive limitations) are not mutually exclusive and may both be relevant factors in receiver decision making.³¹

Irrespective of the source of the high-type behavior among subjects, our experimental design significantly varies the overall share of high-type receivers. This relates our paper to a broader literature on social capital and its impact on credit.³² Our *bl*-, *hr*-, and *le*-treatments can be viewed as “high-social capital” environments because, in addition to a computerized receiver that always returns, a nontrivial fraction of receivers exhibits high-type behavior. In contrast, the *vcle*-treatments can be viewed as a “low-social capital” environment due to the presence of three (computerized) untrustworthy receivers. With this characterization in mind, our experiment suggests that gradualism is redundant and sometimes even costly in

³¹ Prosociality and cognitive limitations may also be relevant in explaining why we see some difference in the share of high-type behavior across the gradual game and its binary version. If prosociality is the main reason behind the high-type behavior, then the difference may be due to some form of “crowding out”. That is, the frequent use of testing strategies in the gradual game may signal that the norm in that game is to default and that the sender does not trust the receiver’s intrinsic motivation to return. As a result, prosocial receiver behavior may be crowded out in the gradual game (see Bohnet et al. (2001) and Falk and Kosfeld (2006) as examples for crowding out of prosocial behavior, and Gneezy et al. (2011) for an extensive survey of the literature). If cognitive limitations are the main reason behind the high-type behavior, then the difference may be due to the fact that the gradual game makes senders more strategic and enables them to understand the monetary incentives better.

³² As an example, Fukuyama (1995) defines social capital as “the existence of a certain set of informal values or norms shared among members of a group that permit cooperation among them.”

high-social capital environments, but beneficial in an environment characterized by low social capital. This interpretation finds support, if indirectly, from various results reported in the literature. For example, Guiso et al. (2004) measuring social capital across Italy report that in “social-capital-intensive areas, households are also more likely [...] to obtain credit when they demand it” and that “the likelihood of receiving a loan from a relative or a close friend decreases with the level of social capital that prevails in the area.” Karlan et al. (2009) develop and find empirical support for a model in which the social network influences how much individuals trust each other. Their results indicate that a dense social network generates high (bonding) social capital, which is particularly important for access to increased (informal) borrowing. Moreover, perhaps reflecting different levels of social capital among native Spaniards and between native Spaniards and immigrants, Diaz-Serrano and Raya (2014) “observe that immigrant borrowers are charged substantially higher interest rates in their mortgages than their native counterparts” and add that these “differentials remain high and significant even after controlling for differences in creditworthiness and other factors.” Arguably, these are examples of what our experimental results suggest, namely, that in a high (low) social capital environment a binary (gradual) “institution of trust” will be more efficient or emerge endogenously more often.

Our interpretation is seemingly also supported by the prevalence of progressive lending in microcredit. Still, we are not aware of any field study showing the optimality of the steep progressive-lending schemes employed in microfinance, and this may be an important direction for research in the field especially relevant for non-profit lenders. Another possible direction for future research would be to implement gradual and binary games in the field to see whether our findings about the costs and benefits of gradualism also emerge in naturally occurring low- and high-social capital environments, such as within and across groups of people with different ethnic backgrounds or religious beliefs.

Appendix. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.geb.2021.07.008>.

References

- Anderhub, V., Engelmann, D., Güth, W., 2002. An experimental study of the repeated trust game with incomplete information. *J. Econ. Behav. Organ.* 46, 197–216.
- Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. *Econ. J.* 103 (418), 570–585.
- Aoyagi, M., Fréchette, G.R., 2009. Collusion as public monitoring becomes noisy: experimental evidence. *J. Econ. Theory* 144 (3), 1135–1165.
- Baker, G., Gibbons, R., Murphy, K.J., 2002. Relational contracts and the theory of the firm. *Q. J. Econ.* 117 (1), 39–84.
- Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10 (1), 122–142.
- Bernard, M., Fanning, J., Yuksel, S., 2018. Finding cooperators: sorting through repeated interaction. *J. Econ. Behav. Organ.* 147, 76–94.
- Bohnet, I., Frey, B.S., Huck, S., 2001. More order with less law: on contract enforcement, trust, and crowding. *Am. Polit. Sci. Rev.* 95, 131–144.
- Boyd, R., Lorberbaum, J.P., 1987. No pure strategy is evolutionary stable in the repeated prisoner's dilemma. *Nature* 327, 58–59.
- Brandts, J., Figueras, N., 2003. An exploration of reputation formation in experimental games. *J. Econ. Behav. Organ.* 50 (1), 89–115.
- Brown, M., Serra-Garcia, M., 2017. The threat of exclusion and implicit contracting. *Manag. Sci.* 63 (12), 4081–4100.
- Brown, M., Falk, A., Fehr, E., 2004. Relational contracts and the nature of market interactions. *Econometrica* 72, 747–780.
- Bull, C., 1987. The existence of self-enforcing implicit contracts. *Q. J. Econ.* 102 (1), 147–159.
- Camera, C., Casari, M., 2009. Cooperation among strangers under the shadow of the future. *Am. Econ. Rev.* 99 (3), 979–1005.
- Camerer, C., Weigelt, K., 1988. Experimental tests of a sequential equilibrium reputation model. *Econometrica* 56, 1–36.
- Cochard, F., Vana, P.N., Willinger, M., 2004. Trusting behavior in a repeated investment game. *J. Econ. Behav. Organ.* 55, 31–44.
- Dal Bó, P., 2005. Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *Am. Econ. Rev.* 95 (5), 1591–1604.
- Dal Bó, P., Fréchette, G.R., 2011. The evolution of cooperation in infinitely repeated games: experimental evidence. *Am. Econ. Rev.* 101, 411–429.
- Dal Bó, P., Fréchette, G.R., 2019. Strategy choice in the infinitely repeated prisoners' dilemma. *Am. Econ. Rev.* 109 (11), 3929–3952.
- Diaz-Serrano, L., Raya, J.M., 2014. Mortgages, immigrants and discrimination: an analysis of the interest rates in Spain. *Reg. Sci. Urban Econ.* 45, 22–32.
- Duffy, J., Ochs, J., 2009. Cooperative behavior and the frequency of social interaction. *Games Econ. Behav.* 66 (2), 785–812.
- Embrey, M., Fréchette, G.R., Lehrer, S.F., 2015. Bargaining and reputation: an experiment on bargaining in the presence of behavioural types. *Rev. Econ. Stud.* 82 (2), 608–631.
- Engle-Warnick, J., Slonim, R., 2004. The evolution of strategies in a trust game. *J. Econ. Behav. Organ.* 55, 553–573.
- Engle-Warnick, J., Slonim, R., 2006a. Learning to trust in indefinitely repeated games. *Games Econ. Behav.* 54, 95–114.
- Engle-Warnick, J., Slonim, R., 2006b. Inferring repeated game strategies from actions: evidence from trust game experiments. *Econ. Theory* 28, 603–632.
- Falk, A., Kosfeld, M., 2006. The hidden costs of control. *Am. Econ. Rev.* 96 (5), 1611–1630.
- Fischbacher, Urs, 2007. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10 (2), 171–178.
- Fréchette, G.R., Yuksel, S., 2017. Infinitely repeated games in the laboratory: four perspectives on discounting and random termination. *Exp. Econ.* 20, 279–308.
- Fudenberg, D., Dreber, A., Rand, D.G., 2012. Slow to anger and fast to forgive: cooperation in an uncertain world. *Am. Econ. Rev.* 102, 720–749.
- Fukuyama, F., 1995. *Trust: The Social Virtues and the Creation of Prosperity: The New Foundations of Global Prosperity*. The Free Press, New York.
- Ghosh, P., Ray, D., 1996. Cooperation in community interaction without information flows. *Rev. Econ. Stud.* 63, 491–519.
- Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *J. Econ. Perspect.* 25 (4), 191–209.
- Grosskopf, B., Sarin, R., 2010. Is reputation good or bad? An experiment. *Am. Econ. Rev.* 100 (5), 2187–2204.
- Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development. *Am. Econ. Rev.* 94, 526–556.
- Halac, M., 2012. Relational contracts and value of relationships. *Am. Econ. Rev.* 102 (2), 750–779.
- Huck, S., Lünser, G.S., Tyran, J.-R., 2016. Price competition and reputation in markets for experience goods: an experimental study. *Rand J. Econ.* 47 (1), 99–117.
- Karlan, D., Mobius, M., Rosenblat, T., Zeid, A., 2009. Trust and social collateral. *Q. J. Econ.* 124 (3), 1307–1361.
- Kartal, M., 2018. Honest equilibria in reputation games: the role of time preferences. *Am. Econ. J. Microecon.* 10 (1), 278–314.
- Kartal, M., Müller, W., 2018. A new approach to the analysis of cooperation under the shadow of the future: theory and experimental evidence. Available at SSRN: <https://ssrn.com/abstract=3222964>.

- Kranton, R., 1996. The formation of cooperative relationships. *J. Law Econ. Organ.* 12, 214–233.
- Kreps, D., Wilson, R., 1982. Reputation and imperfect information. *J. Econ. Theory* 27 (2), 253–279.
- Kreps, D., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoners' dilemma. *J. Econ. Theory* 27 (2), 245–252.
- Levin, J., 2003. Relational incentive contracts. *Am. Econ. Rev.* 93 (3), 835–857.
- MacLeod, W. Bentley, Malcomson, J.M., 1989. Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica* 57, 447–480.
- McMillan, J., Woodruff, C., 1999. Interfirm relationships and informal credit in Vietnam. *Q. J. Econ.* 114 (4), 1285–1320.
- Milgrom, P., Roberts, J., 1982. Predation, reputation, and entry deterrence. *J. Econ. Theory* 27 (2), 280–312.
- Morduch, J., 1999. The microfinance promise. *J. Econ. Lit.* 37 (4), 1569–1614.
- Neral, J., Ochs, J., 1992. The sequential equilibrium theory of reputation building: a further test. *Econometrica* 60, 1151–1169.
- Palfrey, T.R., Rosenthal, H., 1994. Repeated play, cooperation and coordination: an experimental study. *Rev. Econ. Stud.* 61 (3), 545–565.
- Rauch, J., Watson, J., 2003. Starting small in an unfamiliar environment. *Int. J. Ind. Organ.* 21 (7), 1021–1042.
- Sobel, J., 1985. A theory of credibility. *Rev. Econ. Stud.* 52 (4), 557–573.
- Vespa, E., Wilson, A.J., 2019. Experimenting with the transition rule in dynamic games. *Quant. Econ.* 10 (4), 1825–1849.
- Watson, J., 1999. Starting small and renegotiation. *J. Econ. Theory* 85, 52–90.
- Watson, J., 2002. Starting small and commitment. *Games Econ. Behav.* 38, 176–199.
- Wilson, A.J., Vespa, E., 2020. Information transmission under the shadow of the future: an experiment. *Am. Econ. J. Microecon.* 12 (4), 75–98.