Games and Economic Behavior 102 (2017) 624–644

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Equilibrium trust ☆

Luca Anderlini [a,*], Daniele Terlizzese [b]

[a] Georgetown University, United States
[b] EIEF and Bank of Italy, Italy

ARTICLE INFO

Article history:
Received 5 November 2015
Available online 9 March 2017

JEL classification:
D80
D89
C79

Keywords:
Trust
Social norms
Multiple equilibria
Enforcement

ABSTRACT

Trusting beliefs can be exploited. A trustful player who is cheated too often, should start trusting less, until her beliefs are correct. For this reason we model trust as an *equilibrium* phenomenon. Receivers of an offer to transact choose whether or not to cheat. Cheating entails a cost, with an idiosyncratic component and a socially determined one, decreasing with the mass of players who cheat. The model either has a unique equilibrium level of trust (the proportion of transactions not cheated on), or two — one with high and one with low trust. Differences in trust can result from different fundamentals or from different equilibria being realized. Surprisingly, under certain conditions these two alternatives are partially identifiable from an empirical point of view. Our model can be reinterpreted with the cost of cheating arising from an enforcement mechanism that punishes cheaters in a targeted way using limited resources.

1. Introduction

1.1. Motivation and overview

Not many transactions are carried out using rotating security trays with money on one side and the purchased good on the other. Yet, without these devices or some equivalent arrangement, the standard selfish players that populate modern economic models would be unable to trade, barring repeated interaction or binding contracts. It is instead self-evident that trade among players flourishes way beyond what this hypothetical world would look like.

The lubricant that makes so many transactions take place is *trust*. This is the belief that economic players hold that the other side of the transaction will not behave in a completely opportunistic way, thus impeding mutually advantageous exchange.[1]

Our purpose here is to build a simple model of trust as an *equilibrium* phenomenon. The players' beliefs about not being cheated — their level of trust as we just defined it — should be endogenously determined in equilibrium, and hence *correct*.

☆ We are grateful to Helena Aten, Larry Blume, Martin Duwfenberg, Hülia Eraslan, Leonardo Felli, Dino Gerardi, Hugo Hopenhayn, Guido Kuersteiner, Roger Lagunoff, George Mailath, Facundo Piguillem, Aleh Tsyvinski, Bill Zame and seminar participants for helpful comments. We also thank three anonymous referees and the Advisory Editor of this Journal for constructive criticism. Luca Anderlini thanks EIEF for their generous hospitality. All errors remain our own.

* Corresponding author at: Department of Economics, Georgetown University, 37th and O Streets, Washington, DC 20007, USA.
E-mail address: luca@anderlini.net (L. Anderlini).

[1] We define trust as a belief, in line with Charness and Dufwenberg (2006), rather than adopting the behavioral definition advocated by Fehr (2009). Since in equilibrium there will be a one-to-one mapping between beliefs and behavior, the distinction for us is largely immaterial. For ease of exposition, we postpone until Subsection 1.2 a discussion of specific contributions from the large literature on the sources and effects of the presence of trust.

http://dx.doi.org/10.1016/j.geb.2017.02.018
0899-8256/

We focus on equilibrium beliefs because of a basic tension between "trusting beliefs", and consequent "trusting behavior", and the incentives to cheat of other players in society. Trusting beliefs can be exploited. However, a trustful player should not be cheated too much; if she is, she would change her belief and start trusting less. This tension drives our interest for what an equilibrium model of trust can generate.

We are not the first to stress the need for an equilibrium analysis of trust. Within the framework of psychological games, Huang and Wu (1994) and Dufwenberg (1996, 2002) analyze equilibrium trust in models that bear many resemblances to ours. What is new in this paper is an analysis of both the extensive and intensive margins of trusting behavior (not only whether to trust but also how much) and elements of heterogeneity (not all agents deserve to be trusted equally).[2] The conjunction of these two features allows us to explore novel comparative statics which we think shed light on the role of multiple vs unique trust equilibria in the interpretation of empirical evidence.

To model trust as an equilibrium phenomenon we place a "cost of cheating" in the players' utility function. The cost of cheating has two components. One which is an exogenous, idiosyncratic characteristic of each player, and another which is socially determined by the behavior of others. The less common cheating behavior is in society, the higher is the cost of cheating for individual players. This feedback component of the cost of cheating is a central ingredient of our analysis. We interpret it as reflecting a social norm, determined by the average behavior of others. It can also be given a somewhat different psychological underpinning, in the context of the theory of psychological games (we postpone a discussion of this point to Subsection 1.2). The experimental evidence presented in Charness and Dufwenberg (2006) offers strong empirical support for the presence of a social feedback of a similar nature. Alternatively, we might think of the feedback as resulting from an enforcement technology, whose effectiveness depends, for given resources, on the average behavior. We return to this alternative interpretation in Section 6.

Our model is deliberately kept simple in the extreme. However, as we will argue below, the flavor of our results is independent of many of the stark features of our model. Two features of our set-up are worth mentioning at the outset.

First, our model is *not* dynamic, nor should it be interpreted as a "reduced form" of a dynamic set up. Repeated interactions can and have been used successfully to generate cooperative and trust-like behavior, together with a very large variety of other equilibria. However, both in real life and in laboratory experiments trust seems to emerge even in one-shot interactions, in situations that would seem to call for "swivel-tray trading" if trust were not present. To sharpen our understanding of these situations, we work with a model that avoids reputational issues and more generally repeated interaction altogether.

Secondly, our set up is a "one-sided" model of trust, with one player making a proposal, and a responder who can decide whether to cheat or not. Admittedly, trust is often required on both sides of a transaction, and in those cases trust should be modeled as a two-sided phenomenon in which both players in a match have a choice of whether to cooperate or not. Our approach is therefore restrictive. There is, however, a large and growing experimental literature on trust games and reciprocity that explores the emergence of trust in situations akin to the one we model.

In spite of its extreme simplicity, we believe our set up provides a rich enough framework to address the well documented diversity of levels of trust in different societies. Indeed, depending on the configuration of preferences and other parameters, our model either generates a unique equilibrium (with a single equilibrium level of trust), or two equilibria, one with a higher and another with a lower level of trust. Moreover, by varying the parameters of the model, higher or lower levels of equilibrium trust can be obtained without switching across different equilibria. Given this rich set of possible equilibrium outcomes, the model allows us to frame in a natural way what seems a key question concerning the levels of trust in different societies. When we observe different levels of trust across different societies, is this due to a difference in the *fundamental parameters* that underpin the different societies, or is it possible that the *same fundamental parameters* give rise to *different equilibria*? Obviously these two possibilities have vastly different policy implications, and it is therefore important to have a framework in which they can be made precise, and hence disentangled.

Of course, unless the multiplicity can be somehow validated empirically, models with multiple equilibria are no more than a theoretical benchmark providing an — albeit important — only *potential* interpretation of reality. A key insight from our model is that the two cases mentioned above, under some conditions, are *partially identifiable*: if different levels of trust result from multiple equilibria, then the level of trust must be negatively correlated with the size of individual transactions. A positive correlation can only emerge if different levels of trust were to result from differences in the parameters of the model.[3] To our knowledge, the possibility to empirically disentangle multiple vs. unique equilibrium regimes is new.

As we mentioned above, our model can be reinterpreted so that the social feedback component of the cost of cheating comes instead from an enforcement technology that punishes cheating using limited resources in a targeted way. We find that, in the multiple equilibria regime, an infinitesimal increase in the resources devoted to enforcement can yield a discontinuous increase in the level of total activity in the economy. In the single equilibrium regime, instead, the level of activity changes continuously with the resources spent in the enforcement technology.

[2] Heterogeneity is also considered by Attanasi et al. (2016). See footnote 8.

[3] Since the term "partial identification" has been used before Phillips (1989), it is useful to be precise as to the meaning we give it here. The identification is *partial* in the sense that we cannot rule out that two equilibria, corresponding to two sets of parameters, entail a negative correlation between trust levels and size of individual transactions. Therefore, while a positive correlation excludes that the different levels of trust results from multiple equilibria with unchanged parameters, the observation of a negative correlation is inconclusive.

1.2. Related literature

The starting point of this paper — an element of trust is needed for most transactions to take place — is uncontroversial. Long ago Arrow (1972, p. 357) noted that "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time." In the absence of "instantaneous exchange," an element of trust is required.

Arrow based his comment on the path-breaking study by Banfield (1958) of the devastating effects of the lack of trust on a "backwards" small community in southern Italy pervaded by "amoral familism." Following Banfield (1958) a large literature has blossomed on the roots and effects of the lack, or presence, of trust.

The literature is way too large and varied to attempt even a reasoned outline here, let alone a survey. We only recall how Putnam (1993) documents the heterogeneous levels of "social capital" in different regions of Italy and its role in fostering growth. A couple of years later, Fukuyama (1995) published an influential monograph concerning the positive role of trust in large firms and hence economic growth.[4] Bohnet et al. (2010) document crucial differences in trusting behavior across Gulf and Western countries. More recently Bigoni et al. (2016) find that the differences in cooperative behavior between northern and southern Italy cannot be explained by proxies for either social capital or amoral familism. They conclude that persistent differences in social norms are responsible for the observed gap. We believe that our set-up generating multiple equilibria with different levers of trust resonates well with their empirical findings and analysis.

We also selectively recall the contributions by Knack and Keefer (1997), La Porta et al. (1997), and more recently Guiso et al. (2004). These studies all document in a variety of ways how the presence of trust is correlated with desirable economic outcomes.

Our one-sided set-up is closely related to the study of trust in experiments with a sender and a receiver that goes back to Fehr et al. (1993) and Berg et al. (1995) and is more recently found in Glaeser et al. (2000), Irlenbusch (2006), Sapienza et al. (2013) and Butler et al. (2016) among others. One way to place our contribution in the extant literature is that we examine the equilibrium properties of a canonical version of these experimental games.

The social component of our "cost of cheating" is akin to the psychological cost which, in the spirit of the psychological games first introduced by Genakoplos et al. (1989), was assumed by Huang and Wu (1994) and by Dufwenberg (1996, 2002) to analyze the equilibria of a trust game between a trustor (what we call in this paper the offerer) and a trustee (the receiver, in our setting). In those papers the trustee who cheats incurs a payoff penalty proportional to his expectation of the trustor's expectation of the probability that the trustee does not cheat (in Dufwenberg, 1996, 2002) or of the proportion of the trustees who do not cheat (in Huang and Wu, 1994).[5] The idea is that the trustee experiences a belief-based, psychological cost if he cheats when the trustor does not expect him to do so. This might sustain an equilibrium in which the trustee does not cheat.

Since in equilibrium beliefs are correct, the role of higher order beliefs in shaping the cost, postulated in these papers, boils down to including in the cost a component proportional to the equilibrium share of trustee who do not cheat (the equilibrium trust, in our setting), which is precisely what we do in this paper. In fact, the equilibrium of our model is isomorphic to the equilibrium of a two-stage psychological Bayesian game with heterogeneous receivers, whose utility has both an idiosyncratic and a psychological component, the latter dependent on the offerer belief about the receiver's action (or, equivalently in equilibrium, on the receiver second-order belief about his own action).[6] There is, however, a subtle difference between interpreting the cost as resulting from noncompliance with a social norm or from betrayal of the expectations of a well identified partner. We believe that the former is more in line with the one-shot, anonymous nature of the interaction which we emphasize in this paper.

Dufwenberg (2002) interpreted the psychological component of the cost as reflecting "guilt aversion." In later work (Battigalli and Dufwenberg, 2007, 2009), guilt aversion has been given a more precise meaning, based on the difference between the outcome the offerer was expecting to obtain and the actual outcome.[7] To be interpretable as reflecting this more specific notion of guilt aversion, our model would need to include in the cost also a term depending on the size of the transaction, since the expected outcome for the offerer is given by the product of the probability that the receiver does not cheat and the size of the transaction. Indeed, making the cost of cheating sensitive to the size of the transaction to be cheated seems a natural and interesting extension of our baseline assumption, which might be consistent also with other psychological justifications (for example, reciprocity, as in Rabin, 1993 and Dufwenberg and Kirchsteiger, 2004).

We explore this issue and the robustness of our results to alternative formulations of the cost of cheating in Section 5. There, we show that our qualitative results continue to hold, as long as the cost of cheating increases with the size of the

[4] The literature documenting the impact of "social capital" or "trust" on income, wealth and growth rates also includes some notable skeptics. To our knowledge, the most prominent one is Solow (1995), who in turn cites evidence from Kim and Lau (1994) and Young (1994). For a survey of much of the literature on trust and social capital we cite here, including an account of the debate we have just mentioned, see Sobel (2002).

[5] Huang and Wu (1994) [p. 393] wrote: "the more prevalent corruption is, the less intense is the remorse suffered from corrupt behavior; and conversely, the less corruption there is, the more regret from violating a social norm not to be corrupt." This is precisely our interpretation of the social component of the cost.

[6] We are grateful to an anonymous referee for pointing this out.

[7] In Huang and Wu (1994) and Dufwenberg (1996, 2002) the decision by the trustee is binary (whether to cheat or not), so the belief about his action and the belief about the outcome resulting from his action are proportional. Therefore the looser and the more precise notions of guilt aversion are in fact equivalent.

transaction in a less than proportional way. This, we believe, is fairly intuitive. The tension we are exploring in this paper is one between the benefit of increasing the scale of the transactions, which requires trust, and the incentive to cheat, which increases with that scale. In our set-up, the benefit is assumed to increase proportionally with the size of the transaction. As long as the cost of cheating does not increase more than proportionally with that size, the tension remains and our qualitative results are confirmed. If instead the cost is assumed to increase more than proportionally, the tension eventually disappears and the set of equilibria of the game changes qualitatively: when the upper bound to the size of the transaction is large enough, the only equilibrium is one of full trust (no cheating) and maximal transaction size. A similar outcome is obtained assuming that the incentive to cheat is kept in check by reciprocity. There again, a large enough offered transaction is interpreted as a kind act by the offerer, to which the receiver wants to respond kindly. Therefore, a large offer is no longer a temptation to cheat, and the tension evaporates. These results clarify that the scope of our analysis is limited to those cases in which a conflict between the gains from trusting a transaction partner and the risk of being taken advantage of remains in place. Our set-up is not suited to analyze those cases in which the interests of the two parties become aligned.

By simply assuming a cost component proportional to the equilibrium trust, instead of explicitly adopting a psychological game approach, we gain in tractability and we can explore aspects of the problem so far neglected. In particular, our offerer has a continuous choice variable as opposed to the binary one in Huang and Wu (1994) and Dufwenberg (1996, 2002), and this is essential to study both extensive and intensive margins of the trusting behavior. Moreover, our set-up allows for a simple treatment of heterogeneity among the trustees, another aspect which is clearly relevant to understand the role of trust and which is technically much more challenging in the context of a full blown psychological game.[8] Heterogeneous cheating costs yield, in a simple way, an equilibrium in pure strategies with an interior level of trust (neither full nor zero trust), while psychological games with no heterogeneity only obtain interior levels of trust in a mixed strategy equilibrium.

Admittedly, our approach has a cost. By cutting through the chain of higher order beliefs we cannot explore the forward induction arguments which Dufwenberg (1996, 2002) and Battigalli and Dufwenberg (2009) have used to select among the possible equilibria. But while these arguments are a powerful tool in the case of a single offerer and a single receiver, we pursue a model in which offerers and receivers are anonymously matched after being drawn from large populations. We believe that a forward induction argument loses some of its appeal in this case.[9]

Charness and Dufwenberg (2006) present experimental evidence showing that trustees who guess that the trustors' average guess about the proportion of trustees not cheating is higher are less likely to cheat. Although their experimental setting only explores best-response behavior, and cannot be compared directly to our equilibrium set-up, their results empirically confirm the presence of a social feedback in the cost of cheating.[10] Fehr (2009) also discusses at length the possible multiple sources of trusting behavior, identifying social feedbacks (though his analysis focuses on feedbacks affecting the offerer, while we are here concerned with the receiver).

Feddersen and Sandroni (2006) explore the effects of "ethical" social feed-back mechanisms not unlike the one we consider here, and their impact on the equilibria of voting models.[11] Horst and Scheinkman (2006) are concerned with the general theoretical problems of models with social feedback variables, particularly with the (far from trivial) issues that arise in proving the existence of equilibrium in general in this class of models.

Dixit (2003) and Tabellini (2008) are both theoretical contributions to the literature on trust.[12] Their main focus is on the differential effects of distance and society's size on the sustainability of trust. In both cases, trust is modeled as a dynamic two-sided phenomenon with repeated interaction between players. In our model, play only takes place once, and trusting and trustworthy equilibrium behavior can be traced directly back to our preferences embodying the social feed-back we have described above.

Finally, the theoretical literature on repeated interactions, from which we purposedly stay away, is also vast. We simply refer the interested reader to the monograph by Mailath and Samuelson (2006), which also has a comprehensive list of references.

1.3. Plan of the paper

The rest of the paper is structured as follows. In Section 2 we describe the basic model in detail and make precise what constitutes an equilibrium in our set up. In Section 3 we characterize the set of possible equilibria of the model. In Section 4 we highlight how high and low trust equilibria may arise from either differences in the fundamental parameters of the model, or a switch across multiple equilibria supported by the same set of parameter values. In this section we also

[8] A recent working paper by Attanasi et al. (2016), perhaps the first to allow heterogeneity in a psychological game of trust, highlights the technical difficulties of dealing with heterogeneous higher order beliefs.

[9] In Huang and Wu (1994) the trust game considered is, as in our case, one in which the offerer faces a continuum of receivers, and forward induction is not used.

[10] Charness and Dufwenberg (2006) recognize some tension between the theoretical model postulating guilt aversion, which refers to one-on-one pairings between trustor and trustee, and the set up of their experiment, which elicit expectations about averages of pairings. Given that we postulate random pairings among large populations of offerers and receivers, and focus on the feedback from the equilibrium share of (non) cheated transactions, their experimental design seems relevant for our approach, notwithstanding the fact that they stay clear from equilibrium predictions.

[11] In other related works, Dufwenberg et al. (2011) investigate "other-regarding" preferences in general equilibrium models, while Blume (2004) studies the effect of "stigma" in a dynamic model.

[12] See also Dixit (2004).

spell out the identifiable characteristics of high and low trust equilibria in these two cases, and we proceed to characterize aggregate activity levels of the different equilibria. In Section 5 we consider several variants of our baseline cost function, and we assess the robustness of our results to these alternative formulations. Section 6 provides a re-interpretation of the socially generated component of cheating costs as stemming from an enforcement technology that targets and punishes cheaters with limited resources available. Finally, Section 8 briefly concludes.

For ease of exposition, all proofs have been relegated to Appendix A.

2. Set-up

2.1. The model

There are two varieties of players, offering players — or simply $\mathcal{O}$ players — and receiving players — or simply $\mathcal{R}$ players. There is a continuum of $\mathcal{O}$ players of mass 1, and similarly a continuum of $\mathcal{R}$ players of mass 1.

The $\mathcal{O}$ and $\mathcal{R}$ players are then randomly matched to form a unit mass of pairs. The only thing that is of consequence here is that an $\mathcal{O}$ player should not know the "cost of cheating" (to be defined shortly) of the player she is matched with.[13]

Each $\mathcal{O}$ player makes an offer $x \in [0, 1]$ to the $\mathcal{R}$ player in her match. The offer generates a total surplus of $2x$ to be split equally between $\mathcal{O}$ and $\mathcal{R}$ if the transaction goes through without "cheating" on the part of $\mathcal{R}$. So, if $\mathcal{R}$ does not cheat, an offer of x generates a payoff of x for both $\mathcal{O}$ and $\mathcal{R}$.[14]

It is the $\mathcal{R}$ player in the match who decides whether to cheat or not. After receiving an offer x from the $\mathcal{O}$ player in the match, $\mathcal{R}$ may decide to cheat and grab the entire surplus $2x$ instead of abiding by what the splitting procedure suggests. However, if he cheats, $\mathcal{R}$ will also suffer a cost c.[15] Therefore, $\mathcal{R}$ will cheat if $2x - c > x$ or equivalently $x > c$, and will not cheat otherwise.[16]

The total cost of cheating c has two components. One depends on the exogenously given "type" of the $\mathcal{R}$ player and the other is determined by the behavior of other players in the model.

For simplicity, we assume that there are just two types of $\mathcal{R}$ players, "high" (H) and "low" (L).[17] The exogenous component of the cost of cheating is $t_L \in (0, 1)$ for type L and $t_H \in (0, 1)$ for type H, with $t_L < t_H$. The proportion of type H is denoted by $p \in (0, 1)$ throughout.

The component of c that is "socially determined" is the same for all players.[18] Let s be the proportion of $\mathcal{R}$ players who do *not* cheat. We simply set the social component of the cheating cost to equal s and we take the two components of the cheating cost to combine in a linear way.[19] For an $\mathcal{R}$ player of type $\tau \in \{L, H\}$, the total cost of cheating if a proportion s of transactions go through without cheating is given by

$$c(\tau, s, \alpha) = \alpha t_\tau + (1 - \alpha) s \tag{1}$$

with $\alpha \in (0, 1)$ a parameter that measures (inversely) the relative social sensitivity of the players' cost of cheating. Note that assuming that the coefficients of t_τ and s sum to 1 in equation (1) is restrictive, since it fixes the overall sensitivity to total cost. However, this can be viewed as a normalization ensuring that the order of magnitude of x and c is the same.

As we have already remarked, the social component s of c is a critical ingredient of our model. We think of it as embodying the influence of social norms on individual behavior. When fewer people in society cheat, those who do are in some sense further away from the social norm, and this has a "moral cost."

The fact that the right hand side of (1) is linear in its two arguments simplifies our analysis considerably. The actual functional form is largely inessential, but one of its implications is not. In particular, our partial identification result (Proposition 3) below does depend on the fact that the *elasticity* of the total cost of cheating to s should not be "too high". Roughly speaking, if we go from an equilibrium where the L types cheat to another in which they do not, then, for a *given* value of s the equilibrium level of x must be lower to make cheating unprofitable for the L types. However, as the L types change their behavior from cheating to not cheating the equilibrium level of s becomes higher, and this raises the cheating cost for all types. For our partial identification result to hold the first effect must dominate over the second. As it turns out, this

[13] Since we do not consider repeated interaction, the other details of the matching process are completely inessential.

[14] The fact that the surplus is split equally simplifies our calculations, but is inessential.

[15] Note that we are assuming that $\mathcal{R}$ has the choice of whether to cheat or not even when $x = 0$. If he cheats after an offer equal to zero, her payoff will therefore be $-c$. However, when $x = 0$, there is, so to speak, nothing to grab. Hence, an alternative would be to assume that $\mathcal{R}$ does not have a choice of whether to cheat or not when $x = 0$. Proceeding as we do simplifies the analysis but does not impact the results.

[16] Our implicit assumption that when $\mathcal{R}$ is indifferent he will necessarily not cheat simplifies the analysis but is in fact without loss of generality. In equilibrium, the cheating set defined here and formally in (2) below must be open even if it were allowed in principle to be closed. The reason is that if it were not, then the optimal offer of $\mathcal{O}$ players could not be defined because the acceptance set would have to be open, and hence $\mathcal{O}$'s optimal offer could not be well defined.

[17] The overall flavor of our results easily generalizes to an world with any finite number of types. Many of our results also have analogues in a world with a continuum of types. We proceed in this way for the sake of simplicity.

[18] Again, this is to keep things simple. One could imagine heterogeneous cost "sensitivities" to the behavior of others, and this could easily be accommodated in our set up.

[19] We return to the role of this assumption shortly.

requires that the elasticity of $c(\tau, \alpha, s)$ with respect to s be less than one, and the linearity postulated in (1) is sufficient to guarantee that this is the case.[20]

Finally, as we have mentioned already, the effect of s on c can also be re-interpreted as a cost stemming from an enforcement technology. For given resources devoted to enforcement, the probability that a cheater is "caught" increases as fewer people cheat, thus increasing the expected cost of cheating. We pursue this interpretation more formally in Section 6 below.

2.2. Equilibrium definition

A strategy profile $\sigma = (\sigma_{\mathcal{O}}, \sigma_L, \sigma_H)$ assigns a number $\sigma_{\mathcal{O}} \in [0, 1]$ to every $\mathcal{O}$ player — the offer x that she makes — and a cut-off value $\sigma_\tau \in [0, 1]$ to each $\mathcal{R}$ player of type $\tau \in \{L, H\}$, indicating that he will cheat if and only if he receives an offer strictly above σ_τ.[21] Notice that once a profile σ is given, a value of s is also given since the expected proportion of transactions that will not involve cheating is determined directly by σ.[22]

An equilibrium in our model is a strategy profile $\sigma^* = (\sigma^*_{\mathcal{O}}, \sigma^*_L, \sigma^*_H)$ such that the $\mathcal{O}$ players maximize their expected return, given the strategies of the $\mathcal{R}$ players of each type (and their relative weight in the population), and the $\mathcal{R}$ players of each type decide optimally whether to cheat or split whatever offer they received, given the strategies of the other $\mathcal{R}$ players of each type (and their relative weight in the population).

We begin the analysis of the players' maximization problem on the $\mathcal{R}$ side. Fix a σ, and therefore a value of s. Consider an $\mathcal{R}$ player of type $\tau \in \{L, H\}$. He will cheat if and only if the offer x he receives satisfies

$$x > c(\tau, s, \alpha) = \alpha t_\tau + (1 - \alpha)s. \tag{2}$$

For given σ, and hence s, using (2) we can compute the mass of players $\mathcal{R}$ who will *not* cheat on any given offer $x \in [0, 1]$, denoted by $P(x, s)$:

$$P(x, s) = \begin{cases} 0 & \text{if} \quad x \in (c(H, s, \alpha), 1] \\ p & \text{if} \quad x \in (c(L, s, \alpha), c(H, s, \alpha)] \\ 1 & \text{if} \quad x \in [0, c(L, s, \alpha)] \end{cases} \tag{3}$$

For given σ, we can therefore write the expected payoff of an $\mathcal{O}$ player offering x as $x\,P(x, s)$ (recall that offering x, she gets a payoff of x whenever she is not cheated). Hence she will choose an x that solves

$$\max_{x \in [0,1]} x\,P(x, s) \tag{4}$$

The solution to (4) is immediate to characterize. Since $c(L, s, \alpha) < c(H, s, \alpha) < 1$, the solution to (4) depends on the comparison between

$$c(L, s, \alpha) \quad \text{and} \quad p\,c(H, s, \alpha) \tag{5}$$

If $c(L, s, \alpha) > pc(H, s, \alpha)$ then it is uniquely optimal to set $x = c(L, s, \alpha)$ — the largest level that ensures that no $\mathcal{R}$ players cheat. If instead $c(L, s, \alpha) < pc(H, s, \alpha)$ the unique solution is to set $x = c(H, s, \alpha)$ — the largest level that ensures that only $\mathcal{R}$ players of type L cheat. Finally if $c(L, s, \alpha) = pc(H, s, \alpha)$, then an $\mathcal{O}$ player is indifferent between making an offer of $c(L, s, \alpha)$ and an offer of $c(H, s, \alpha)$, and hence both values solve (4).

Intuitively, increasing x increases (in jumps, because of the discrete nature of the types) the probability that the offer will be cheated on. The trade-off between increased payoff conditional on not being cheated on and the increase in the probability of cheating is what determines the optimal behavior of $\mathcal{O}$ players.

Before proceeding further, we introduce a simplifying assumption on the behavior of $\mathcal{O}$ players.

Assumption 1 *(Tie break).* Whenever $c(L, s, \alpha) = p\,c(H, s, \alpha)$, all $\mathcal{O}$ players make an offer of $c(L, s, \alpha)$ which is not cheated on with probability one, rather than an offer of $c(H, s, \alpha)$ which is cheated on by all $\mathcal{R}$ players of type L.

For ease of exposition, from now on, when we say that "x solves (4)" we will mean a solution that complies with the tie-breaking rule posited here.

[20] The details of how this elasticity condition guarantees that the first effect dominates the second are discussed in Subsection 4.2, following the statement of Proposition 5.

[21] For simplicity we are focusing on strategies where all players of a given type, that is $\mathcal{O}$, $\mathcal{R}$ of type L and $\mathcal{R}$ of type H, act in the same way. Moreover, we assume that the strategies of the $\mathcal{R}$ players can be summarized by a cut-off value, albeit in principle the players' cheating responses could be based on more than a simple cut-off value. However, it is easy to show that we can restrict the strategy spaces of $\mathcal{R}$ players in this way without loss of generality, using a standard weak-dominance argument. The cheating set can also be shown to be open without loss of generality in equilibrium — see footnote 16 above.

[22] An alternative interpretation of s, in the framework of psychological games, would see it as the expectation of the $\mathcal{R}$ player of the probability that the $\mathcal{O}$ player assigns to not being cheated when offering x.

Assumption 1 makes the behavior of all $\mathcal{O}$ players uniquely determined for any parameter values and any level of s, substantially simplifying the analysis.

Note that the behavior that Assumption 1 postulates can be interpreted as the result of "lexicographic" risk-aversion of the $\mathcal{O}$ players (added to their basic risk-neutrality). Whenever expected values are equal (and only then), random variables with a lower risk are preferred. In particular, when $c(L,s,\alpha) = pc(H,s,\alpha)$, the sure payoff of $c(L,s,\alpha)$ will be preferred to a random payoff equal to 0 with probability $1-p$ and to $c(H,s,\alpha)$ with probability p.

Assumption 1 simplifies our analysis since it rules out, for parameter configurations supporting multiple equilibria, an equilibrium that is "intermediate" between the two that we will focus on. This equilibrium is "between" the two remaining ones, and its presence would not affect our qualitative conclusions in any way.[23]

We can now provide a working definition of what constitutes an equilibrium in our model. Note that, using Assumption 1, the equilibrium behavior of all $\mathcal{O}$ players is summarized by a single number $x \in [0,1]$ — the solution to (4), which is the offer they all make to the $\mathcal{R}$ player they are each matched with. Of course, in equilibrium it must also be the case that the value of s that appears in (4) is the correct one, as determined by the behavior of the $\mathcal{R}$ players given x. This justifies the following definition of equilibrium.

Definition 1 *(Equilibrium).* An equilibrium is a pair (x,s) such that x solves (4) given s and such that

$$P(x,s) = s \tag{6}$$

It is worth noting that our definition of equilibrium implicitly invokes a notion of sequential rationality, as the $\mathcal{R}$ players react optimally, and are known by the $\mathcal{O}$ players to react optimally, to *any* offer they might receive.[24] Such a notion would correspond to the requirement of subgame perfection, if we were to make the assumption that each $\mathcal{R}$ player observes the offer received by all the other $\mathcal{R}$ players, so that the stage of the game in which $\mathcal{R}$ players act would be a proper subgame. Alternatively, our equilibrium would correspond to a sequential equilibrium of a discretized version of our game (to avoid technical measurability problems). Fleshing out these alternative possibilities formally would however be tedious and would not add anything of substance to our analysis.

Given what we know about the behavior of $\mathcal{O}$ players, it is easy to check that only two possibilities are open for the value of s in any equilibrium. Setting an x such that both types of $\mathcal{R}$ players cheat is clearly never optimal. Hence in equilibrium it must be that either $s = 1$ (and no cheating at all takes place), or $s = p$ (and all $\mathcal{R}$ players of type L cheat, while all those of type H do not).

It follows easily that, in equilibrium, if $s = 1$ then $x = c(L,1,\alpha) = \alpha t_L + 1 - \alpha$, and similarly if $s = p$ then $x = c(H,p,\alpha) = \alpha t_H + (1-\alpha)p$.[25]

At this point, it is useful to crystallize some terminology for future use.

Definition 2 *(NC and LC equilibria).* An equilibrium (x,s) with $s = 1$ and $x = c(L,1,\alpha)$ — in which no $\mathcal{R}$ players cheat — is called a *No Cheating (NC)* Equilibrium. An equilibrium (x,s) with $s = p$ and $x = c(H,p,\alpha)$ — in which $\mathcal{R}$ players of type L cheat — is called a *Low Cheating (LC)* Equilibrium.

3. Equilibrium characterization

3.1. NC equilibrium

We can now work out the conditions under which the model has an NC equilibrium.

By definition in this case the equilibrium value of s is 1 — the probability of cheating is 0. Therefore

$$c(L,1,\alpha) = \alpha t_L + 1 - \alpha \quad \text{and} \quad c(H,1,\alpha) = \alpha t_H + 1 - \alpha \tag{7}$$

We already know that in an NC equilibrium $x = c(L,1,\alpha)$. Therefore, to confirm that $[c(L,1,\alpha),1]$ is an equilibrium, we just need to check that the parameters of the model are such that no $\mathcal{O}$ player has an incentive to deviate unilaterally from offering $x = c(L,1,\alpha)$ (which gives her a payoff of precisely $x = c(L,1,\alpha)$ since no cheating takes place).

Deviating to an offer below $c(L,1,\alpha)$ is never profitable since it yields a lower payoff conditional on no cheating taking place, but obviously cannot decrease any further the probability that cheating occurs. Deviating to an offer above $c(L,1,\alpha)$, and hence accepting that $\mathcal{R}$ players of type L will cheat, can yield at most a payoff of $pc(H,1,\alpha)$. In fact this is what an

[23] It is an asymmetric equilibrium of the game with a continuum of players where players of the same type behave differently. Details are available from the authors on request.

[24] Thus, a strategy profile in which all $\mathcal{R}$ players cheat any offer greater than $x^* < \alpha t_L + 1 - \alpha$ and all $\mathcal{O}$ players send x^* would not be an equilibrium, according to our definition. However it would be a non-subgame perfect, or a non-sequential, Nash equilibrium of the appropriately modified version of our game. We are grateful to an anonymous referee for pointing this out.

[25] Note that while the linearity of $c(\cdot)$ allowed us to easily compute the equilibrium offers, aside from guaranteeing monotonicity, it played no critical role in the argument so far.

$\mathcal{O}$ player gets if she makes the largest offer that $\mathcal{R}$ players of type H will not cheat upon, taking as given the equilibrium value of $s = 1$. Hence, using (7), a necessary and sufficient condition for an NC equilibrium to exist is that

$$\alpha t_L + 1 - \alpha \geq p\,[\alpha t_H + 1 - \alpha] \tag{8}$$

3.2. LC equilibrium

The conditions under which an LC equilibrium exists can be worked out in a parallel way. By definition in this case, $s = p$. Hence

$$c(L, p, \alpha) = \alpha t_L + (1 - \alpha)p \quad \text{and} \quad c(H, p, \alpha) = \alpha t_H + (1 - \alpha)p \tag{9}$$

As we noted above, in equilibrium when $s = p$ it must be that $x = c(H, p, \alpha)$. To ensure that $[c(H, p, \alpha), p]$ is an equilibrium we then need to check that the parameters of the model are such that no $\mathcal{O}$ player has an incentive to deviate unilaterally from offering $x = c(H, p, \alpha)$, which yields her an expected payoff of $pc(H, p, \alpha)$. With a logic that is by now familiar, without loss of generality we can consider only the deviation to offering $x = c(L, p, \alpha) -$ the largest offer that will induce no cheating from either type of $\mathcal{R}$ player, taking as given the equilibrium value $s = p$. This deviation yields a payoff of $c(L, p, \alpha)$. Hence, using (9), a necessary and sufficient condition for the model to have an LC equilibrium is

$$p\,[\alpha t_H + (1 - \alpha)p] \geq \alpha t_L + (1 - \alpha)\,p \tag{10}$$

3.3. Multiple and unique equilibria

It is useful to sum up and sharpen our picture of the possible equilibria of the model as a function of the parameter quadruple (α, p, t_L, t_H). For the sake of simplicity, from now on we restrict attention to quadruples away from the boundary of $[0, 1]^4$, satisfying $t_H > t_L$. This dispenses us from having to consider separately some of the boundary cases which would not add anything of interest to our results.

Proposition 1 *(Equilibrium set). The equilibrium set of the model is guaranteed to be non-empty, and can be of three types. A unique NC equilibrium, denoted as the NCU regime; a unique LC equilibrium, denoted as the LCU regime; two equilibria, one LC and one NC equilibrium, denoted as the LCNC (or multiple equilibria) regime.*

If (8) is satisfied and (10) is not, then we are in the NCU regime. If (10) is satisfied and (8) is not, then we are in the LCU regime. Finally, if (8) and (10) are both satisfied, then we are in the multiple equilibria regime.

Although a formal proof of Proposition 1 does not require much more than using some of the observations we have already made, for the sake of completeness we present one in Appendix A. Note that in the multiple equilibria regime only two equilibria are possible because of our simplifying Assumption 1. As we mentioned above, without it we would get a third "intermediate" equilibrium, in which the proportion of transactions not cheated upon is strictly between p and 1.

The three equilibrium regimes of Proposition 1 are also all robust in the standard sense.

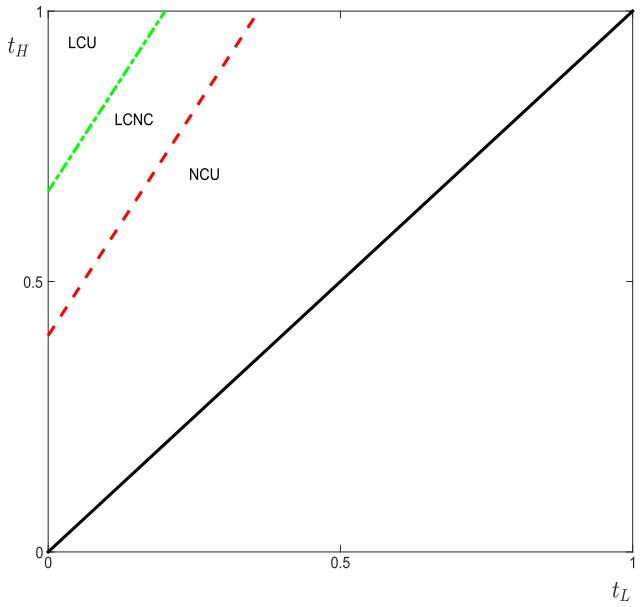
Proposition 2 *(Parametric conditions). The set of parameter quadruples (α, p, t_L, t_H) that yield the NCU regime contains an open set. The same is true for the set of quadruples yielding LCU, and for those yielding LCNC.*

A formal proof of Proposition 2 is in Appendix A. The simple argument behind it hinges on the fact that (8) and (10) can be jointly rewritten as

$$(1 - \alpha)p(1 - p) \leq \alpha(pt_H - t_L) \leq (1 - \alpha)(1 - p) \tag{11}$$

with the first inequality corresponding to (10) and the second to (8). Since $(1 - \alpha)p(1 - p) < (1 - \alpha)(1 - p)$ for all α and p in the interval $(0, 1)$, we are guaranteed to find robust parameter configurations that support each of the three regimes.

The parameter space $\mathcal{P} = \{(\alpha, p, t_L, t_H) \in (0, 1)^4, t_H > t_L\}$ can be partitioned into regions corresponding to the different equilibrium regimes. In order to present an easy to read picture, we fix one value for α and one value for p, and we then divide the set of all relevant t_H and t_L, i.e. the upper diagonal of the unit square, in three regions corresponding to the three regimes: *NCU*, *LCNC* and *LCU*. The figure below is drawn for $\alpha = \frac{1}{2}$ and $p = \frac{3}{5}$. All pairs (t_H, t_L) above the dashed line, representing the locus $\frac{t_L}{p} + \frac{1-\alpha}{\alpha}(1 - p)$, generate *LCU* equilibria. All pairs (t_H, t_L) below the dot-dashed line, representing the locus $\frac{t_L}{p} + \frac{1-\alpha}{\alpha}\frac{(1-p)}{p}$, give rise to *NCU* equilibria. Clearly, all pairs (t_H, t_L) between the dashed and dot-dashed lines yield *LCNC* equilibria. Note that the two lines are parallel and the intercept of the dot-dashed one is always above the intercept of the dashed one (for any choice of α and p in $(0, 1)$).



The figure is drawn for a particular pair of α and p such that all three regions are present; a similar picture would obtain for any pair of α and p such that $\alpha > 1 - p$. With pairs of α and p outside this range, some of the regions would disappear.[26]

We conclude this section with an observation, which we do not fully formalize purely for reasons of space.[27] Although they embody a level of trust s exactly equal to 1, the NCU regime and the NC equilibrium of the NCLC regime are less special than might seem at first sight. The fact that they yield $s = 1$ rather than just a "high" s is an artifact of our choice to consider only two types of $\mathcal{R}$ players, in order to keep the model as simple as possible.

Suppose that a small proportion (say ε) of a third "very low" type were introduced, characterized by an exogenous component of the cost of cheating between 0 and t_L. Then it would be easy to check that the same parameters supporting an NC equilibrium in the two-types model would yield an equilibrium with $s = 1 - \varepsilon$, in which only the "very low" type of $\mathcal{R}$ players cheat. This is clearly possible in a straightforward way whenever the model yields an NC equilibrium which is "strict" in an appropriate sense. With this observation in mind, we will often interpret the NC equilibria as "high trust" equilibria of a general kind, rather than focusing specifically on the fact that they display "full trust".

4. Societies with different levels of trust

4.1. Differences in fundamentals vs. equilibrium switch

Differences of trust levels across different societies are well documented. Moreover, they are often seen to be correlated with (and sometimes held responsible for) phenomena of primary economic importance like income levels and growth rates.

Our simple model highlights that such differences can be traced to two conceptually distinct sources. Two societies may exhibit different levels of trust because they differ in some of their *fundamental parameters* (the vector (α, p, t_L, t_H) in our model). Alternatively, the two societies may, in fact, be the *same* in terms of their *fundamental parameters*, but have *selected different equilibria* among those supported by their common parameter vector. For brevity, we refer to this possibility by saying that their different levels of trust are attributable to a *switch* between different *equilibria*.

In short, we say that an *equilibrium switch* is obtained when the two societies are characterized by the *same* parameter quadruple generating the NCLC regime, the LC equilibrium is realized in one of the societies (the low trust one) while the NC equilibrium is realized in the other (the high trust society).

We instead refer to *differences in fundamentals* when different levels of trust reflect the fact that the two societies are characterized by *different* parameter quadruples, leading to different levels of equilibrium trust. This in turn can happen

[26] For values of α and p such that $\alpha < \frac{1-p}{2-p}$, the intercept of the dashed line (and of course of the dot-dashed line as well) would be above 1. Hence, in this case all pairs (t_H, t_L) would yield an equilibrium of type NCU. For values of α and p such that $\frac{1-p}{2-p} < \alpha < 1 - p$, the intercept of the dot-dashed line would be above 1, while that of the dashed line would be below 1. Hence, in this case all pairs (t_H, t_L) the dashed line would yield an equilibrium of type NCU, all pairs above would give rise to an equilibrium $LCNC$.

[27] The details are available from the authors upon request.

in several different ways, as the two quadruples can be in different regions of the parameter space supporting different equilibrium regimes, or they might support the same equilibrium regime but with different equilibrium trust levels.

Among the many ways in which a difference in fundamentals can arise we will later single out the case in which only the probability mass of the two types of $\mathcal{R}$ players changes, with all other parameters constant. This we refer to as a difference in the *distribution of types*. Specifically, in one society the parameter quadruple is $(\alpha, \overline{p}, t_L, t_H)$, in the other is $(\alpha, \underline{p}, t_L, t_H)$, and in both the LC equilibrium obtains.[28] In the first case the equilibrium trust level is $\bar{s} = \overline{p}$ and in the second it is $\underline{s} = \underline{p} < \bar{s}$.

While it is clear that no two societies will in fact ever be identical in their fundamentals, the theoretical possibility of an equilibrium switch cautions against hastily interpreting different levels of trust as evidence of deep-rooted differences. Moreover, the "partial identifiability" of the equilibrium switch regime that we have mentioned above (and to which we will turn next) provides some guidance to the empirical investigation on the ultimate sources of the observed different levels of trust.

Clearly, the distinction between differences in fundamentals and equilibrium switch has interesting policy implications. Different policy prescriptions will be more appealing according to whether, in order to raise trust levels, a change in fundamentals or an equilibrium switch is needed. Consider a Government seeking to intervene to rectify a low level of trust. If the low trust level is the result of an equilibrium switch, then the Government can attempt to engender a switch to a higher level of trust by re-focusing the *expectations* of the players. Convincing everyone in society that the level of trust is $\bar{s}$ instead of $\underline{s}$ will do the job, since this is capable of becoming a self-fulfilling prophecy.

If instead the low level of trust is due to a difference in fundamentals, then the only way to ensure a high level of trust is to engineer a change in the parameters of the model. This is conceptually and operationally different from a self-fulfilling change in beliefs, and likely harder to achieve, although one might imagine a range of tools that vary from economic incentives to educational programs based on "civic culture" that could be brought to bear.

The conceptual difference between the two hypotheses, difference in fundamentals versus equilibrium switch, might also be of interest to scholars in other disciplines, such as political science or sociology. In essence, a difference in fundamentals points to different norms of behavior being rooted in "anthropological" factors, while an equilibrium switch points in the direction of random events triggering changes that become long-lasting because of self-reinforcement mechanisms at work in society. Our static model is of course silent about the process that might have led to one equilibrium or the other to prevail, but has the merit of verifying that both possibilities are logically consistent and suggests an empirical test that, under some conditions, can identify which of the two hypotheses is true.

4.2. Partial identification: trust and transaction levels

Suppose that we observe two societies, one with a high and the other with a low level of trust. Can we identify whether the differing trust levels are due to an equilibrium switch or to a difference in the fundamentals? As we discussed already in Section 1 and Subsection 2.1, under some conditions the answer is a qualified *yes*.

If the elasticity of the cost of cheating with respect to s is smaller than 1, there are observations that allow us to rule out an equilibrium switch, while other observations are inconclusive. In this specific sense we then say that *partial identification*[29] is possible. The condition on the elasticity is satisfied in the model as specified above, due to the linearity of the cost of cheating in s.

Full identification can be achieved if the range of possible alternatives is further restricted in an appropriate way. In particular, if we know that the only differences in the fundamentals that need to be considered are variations in the distribution of types, then full identification becomes possible.

We begin with the general case, in which all possible variations in parameter quadruples are considered.

The key to identification — partial or full, as the case may be — is the equilibrium level of x, the offer made in equilibrium by all the $\mathcal{O}$ players, which is also the equilibrium individual transaction level.

The following two propositions make our claim precise for the general case.

Proposition 3 *(Individual transaction levels — equilibrium switch). Let a parameter quadruple supporting the NCLC regime be given. Then the level of x in the associated NC equilibrium is* lower *than the level of x in the associated LC equilibrium.*

It follows that if the difference in equilibrium trust between two societies is due to an equilibrium switch *in the sense of Subsection 4.1, then the equilibrium level of x is* negatively correlated *with the equilibrium trust level.*

If we then observe that the equilibrium level of x is higher in the society with higher level of trust we can rule out the possibility of an equilibrium switch.

When it comes to a difference in fundamentals, in principle we should consider any pair of equilibria, of the NC and/or LC type. For completeness, the following proposition examines four cases. The bottom line is that when we allow for

[28] So, clearly, both parameter quadruples must give rise to either the LCU or the NCLC regimes.

[29] See footnote 3 above.

unrestricted differences in fundamentals, the relationship between differences in the equilibrium trust level and the level of individual transaction cannot be pinned down.

Proposition 4 *(Individual transaction levels — change in fundamentals). Fix arbitrarily a parameter quadruple supporting an NC equilibrium. Then we can find parameter quadruples that support LC equilibria with individual transaction levels that can be both higher and lower than in the given NC equilibrium.*

Fix again arbitrarily a parameter quadruple supporting an NC equilibrium. Then we can find parameter quadruples that support NC equilibria with individual transaction levels that can be both higher and lower than in the given NC equilibrium.

Fix arbitrarily a parameter quadruple supporting an LC equilibrium. Then we can find parameter quadruples that support NC equilibria with individual transaction levels that can be both higher and lower than in the given LC equilibrium.

Finally, fix again arbitrarily a parameter quadruple supporting an LC equilibrium. Then we can find parameter quadruples that support LC equilibria displaying a higher level of trust and a lower level of individual transactions than in the original LC equilibrium, or a lower level of trust and a higher level of individual transactions than in the original LC equilibrium.

The picture yielded by Propositions 3 and 4 — namely, partial identification — changes when we consider the case in which parameter differences are restricted to be differences in the distribution of types in the sense mentioned above. In this case the ambiguity highlighted by Proposition 4 no longer holds and a definite conclusion can be reached about the correlation between trust and the size of individual transactions following the difference in parameter values.

Proposition 5 *(Transaction levels — change in distribution of types). Consider two parameter quadruples* $(\alpha, \overline{p}, t_L, t_H)$ *and* $(\alpha, \underline{p}, t_L, t_H)$ *with* $\overline{p} > \underline{p}$*, each giving rise to an LC equilibrium with high* ($s = \overline{p}$) *and low* ($s = \underline{p}$) *trust levels respectively.*

Then the equilibrium level of x associated with $(\alpha, \overline{p}, t_L, t_H)$ *is higher than the equilibrium level of x associated with* $(\alpha, \underline{p}, t_L, t_H)$.

It follows that if the difference in equilibrium trust between two societies is due to a difference in fundamentals in the narrower sense of a difference in the distribution of types*, then the equilibrium level of x is* positively correlated *with the equilibrium trust level.*[30]

Formal proofs of Propositions 3, 4 and 5, which, again, consist of fairly simple algebra, appear in Appendix A. Here, we elaborate on the intuition behind our results, beginning with Proposition 4.

Recall that in any NC equilibrium the level of individual transaction is $\alpha t_L + 1 - \alpha$ while in any LC equilibrium it is given by $\alpha t_H + (1-\alpha)p$. Proposition 4 is then the result of the following observations. If we are allowed to vary all the parameters at will, the necessary and sufficient conditions (8) for an NC equilibrium to exist are compatible with any value of $\alpha t_L + 1 - \alpha$ in $(0, 1)$. Similarly, if we are allowed to vary all the parameters at will, the necessary and sufficient conditions (10) for an LC equilibrium to exist are compatible with any value of $\alpha t_H + (1-\alpha)p$ in $(0, 1)$.

The statement of Proposition 5 is an immediate consequence of the fact that in any LC equilibrium the level of individual transactions is given by $\alpha t_H + (1-\alpha)p$. Hence, since all other parameters are kept constant, it must increase as p increases.

The intuition behind Proposition 3 requires some intermediate steps. Recall that in this case we are concerned with a *single* parameter quadruple supporting the NCLC regime. Call the level of individual transactions in the LC equilibrium $x_{LC} = \alpha t_H + (1-\alpha)p$. In the LC equilibrium an $\mathcal{O}$ player gets an expected payoff of $p\ x_{LC}$, since with probability $1-p$ she is cheated by the $\mathcal{R}$ player she meets and ends up with a payoff of zero. She could however deviate and make the largest offer that will induce *no* cheating from any type of $\mathcal{R}$ players, $\alpha t_L + (1-\alpha)p$. Denote this by x^D_{LC}. By deviating, she would get a payoff equal to her offer for sure, hence her incentive constraint allows us to conclude that $p\ x_{LC} \geq x^D_{LC}$ must be true.

Now consider the NC equilibrium associated with the given parameter quadruple. The equilibrium level of individual transactions now is $x_{NC} = \alpha t_L + 1 - \alpha$, since this is the largest offer that will keep *all* $\mathcal{R}$ players from cheating. The only difference between x^D_{LC} and x_{NC} is in fact given by the higher level of equilibrium trust reflected in x_{NC}. This makes it immediate to check that we have $x^D_{LC} > p\, x_{NC}$.

The two steps we have outlined give that $p\, x_{LC} \geq x^D_{LC}$ and $x^D_{LC} > p\, x_{NC}$ respectively, which together immediately yield the claim of Proposition 3, namely $x_{LC} > x_{NC}$.

To see the critical role of the elasticity of the cost function mentioned above, take a general functional form for the cost of cheating $c(t, s)$. It must still be true that in the low trust equilibrium the offer is the largest one that induces type H players not to cheat — namely $c(t_H, p)$ — and in the high trust (no cheating) equilibrium the offer is the largest one that induces type L players not to cheat — namely $c(t_L, 1)$.[31] The incentive constraint of $\mathcal{O}$ players in the low trust equilibrium requires that what they expect to obtain — namely $pc(t_H, p)$ — be at least as large than what they can get deviating unilaterally (i.e., not changing the equilibrium level of trust) and lowering their offer, so as to induce also the more cheating-prone players (type L) not to cheat, and at the same time raising the probability of completing the transaction (in fact, making sure that the transaction is carried out). Formally, it must then be the case that $pc(t_H, p) \geq c(t_L, p)$. Compare now $c(t_L, p)$

[30] Note that once we fix all parameters except for p, all NC equilibria are the same. Hence, since the effect of a switch from LC to NC is already characterized by Proposition 3, the only relevant comparison is the one treated here — two LC equilibria. All the other cases we treated in Proposition 4 can be ignored.

[31] In other words, in equilibrium the lowest type that does not cheat is always indifferent between cheating and not cheating.

with $pc(t_L, 1)$. If the elasticity of $c(t, s)$ with respect to its second argument is smaller than 1, the increase from $c(t_L, p)$ to $c(t_L, 1)$ is more than compensated by the fall from $c(t_L, 1)$ to $pc(t_L, 1)$. Hence $c(t_L, p) > pc(t_L, 1)$, which together with the incentive constraint yields the desired result — namely $c(t_H, p) > c(t_L, 1)$.

Another way to see why the elasticity cannot be too large for our result to hold is as follows. In the high trust equilibrium the offer must be low enough so that all receiver types prefer not to cheat. In the low trust equilibrium, by contrast, only the type H players do not cheat. Since their cost of cheating is higher, the equilibrium offer can be correspondingly higher.

However, as we go from a high trust equilibrium to a low trust equilibrium, everyone's cost of cheating is lower, which would suggest a lower equilibrium offer in the low trust equilibrium. The elasticity condition we identified above ensures that the first effect dominates the second one.

4.3. Trust and aggregate transactions

One of the novel results of Subsection 4.2 is that, under certain conditions, the level of trust and transaction level x are negatively correlated in the case of an equilibrium switch.

On the other hand, a recurrent theme in the extant literature (see Subsection 1.2) is that of a positive relationship between trust and measures of aggregate income or activity. Does it then follow that these findings exclude switches across multiple equilibria as the root of different trust levels in the societies that have been examined? The answer depends on what we take to be an appropriate measure of aggregate income or activity in our extremely simplified model. We examine two alternatives.

To begin with, it is clear that the equilibrium size of individual transactions is not a good candidate for an aggregate measure of activity of a society, simply because it not an economy-wide aggregate, but an individual measure.

The first alternative that we consider is to focus on those transactions that are *not cheated on* as a measure of aggregate income/economic activity. When a transaction is cheated on it does not go through and, as it were, it goes unrecorded. Although this may seem against a literal interpretation of the model, we view unrecorded transactions as excluded from the *measured* aggregate output.[32]

Simple algebra then tells us that in the case of an equilibrium switch the aggregate (measured) economic activity is unambiguously higher in the NC equilibrium than in the LC equilibrium, as the higher level of individual transactions in the LC equilibrium is more than compensated by the lower proportion of transactions being completed. The observation of a positive correlation between trust levels and levels of aggregate activity is not therefore inconsistent with an equilibrium switch.[33]

The second alternative is to say that all matches initially produce an aggregate social payoff of $2x$, but those cheated on have to be reduced by the cheating cost; in other words, we proceed to aggregate payoffs counting as $2x$ those transactions that are not cheated on, and counting as $2x - c$ those transactions that are cheated on. To the extent that the cheating cost is subjectively perceived by the cheater but has no material counterpart, this alternative measure of aggregate activity in our model is in fact closer to a measure of aggregate *welfare* as opposed to measured aggregate economic activity.

In this case, the comparison of aggregate economic activity between LC and NC equilibria of an NCLC regime is ambiguous. There are parameter values that give rise to the NCLC regime in which the LC equilibrium yields higher aggregate output and other ones in which the NC equilibrium does.[34]

This is perhaps not surprising. Recall that our $\mathcal{R}$ players cheat when they find it *optimal* to do so. Hence, more cheating does not automatically reduce aggregate welfare in society.

5. Alternative formulations

As mentioned in the Introduction, a natural generalization of our formulation for the cost of cheating is to assume that cheating on a larger transaction is, everything else equal, more costly than cheating on a smaller one. Keeping for simplicity the linear functional form to aggregate the two components of the total cost (idiosyncratic and social), we still have 4 possibilities, depending on whether the size of the transaction enters the cost through the idiosyncratic or the social component, and whether it enters in a concave or convex way.

We will show that allowing for the cost of cheating to depend on the size of the transaction, whether through the idiosyncratic or the social component of the cost, leaves our results qualitatively unchanged if the dependence is (strictly) concave. In the knife-edge case of a proportional dependence we confirm the three kind of equilibria revealed by our previous analysis, but the multiple equilibria case becomes non-generic (i.e. it is supported by a measure zero set of parameters). When the dependence is convex, the nature of the equilibria changes in a qualitative way: the only possible equilibrium is one with full trust and maximal transaction size (provided that we set the exogenous upper bound on the latter to be large enough).

[32] When a $\mathcal{R}$ player cheats she walks away with $2x$, which is the same total payoff that the two players $\mathcal{O}$ and $\mathcal{R}$ would split equally if no cheating were to take place. In a richer model, transactions that are cheated on could be modeled as in fact producing a smaller aggregate payoff — as is the case in many versions of the trust game.

[33] To streamline the exposition, the formal statement of this claim (and its proof) appears in Appendix A as Proposition A.1.

[34] Again, to streamline the exposition, the formal statement of this claim (and its proof) appears in Appendix A as Proposition A.2.

The same conclusion would obtain assuming that the psychological mechanism keeping in check the temptation to cheat is a concern for reciprocity. The intuition for these results was provided in Section 1.2.

5.1. Idiosyncratic cost of cheating depending on the transaction size

We first suppose that for an $\mathcal{R}$ player of type $\tau \in \{L, H\}$, the total cost of cheating, if a proportion s of transactions go through without cheating, is given by:

$$c(\tau, s, \alpha, x) = \alpha t_\tau x^\gamma + (1-\alpha)s, \tag{12}$$

in which the dependence on x of the (idiosyncratic part of the) cost is a concave function ($0 \leq \gamma \leq 1$). The idea is that cheating on a larger transaction is, everything else equal, more costly, but the additional pang is progressively smaller. For $\gamma = 0$ we get back the formulation adopted in the preceding Sections.

This formulation of the cost function no longer yields (in general) a closed form solution for the decision of the $\mathcal{R}$ player of type $\tau \in \{L, H\}$. He will cheat if and only if

$$2x - (\alpha t_\tau x^\gamma + (1-\alpha)s) > x, \tag{13}$$

i.e. if and only if

$$x > h(\alpha, t_\tau, s), \tag{14}$$

where $h(\alpha, t_\tau, s)$ is the solution to[35]:

$$x = \alpha t_\tau x^\gamma + (1-\alpha)s, \tag{15}$$

whose existence, continuity and differentiability are guaranteed by the implicit function theorem.[36] Clearly, $h(\alpha, t_H, s) > h(\alpha, t_L, s)$, for all $\gamma \in [0, 1]$.[37] Therefore type L will not cheat for offers $x \in [0, h(\alpha, t_L, s)]$, and he will cheat otherwise; type H will not cheat for offers $x \in [0, h(\alpha, t_H, s)]$, and he will cheat otherwise. The mass of players $\mathcal{R}$ who will *not* cheat on any given offer $x \in [0, 1]$, denoted by $P(x, s)$ is therefore:

$$P(x,s) = \begin{cases} 0 & \text{if} \quad x > h(\alpha, t_H, s) \\ p & \text{if} \quad h(\alpha, t_L, s) < x \leq h(\alpha, t_H, s) \\ 1 & \text{if} \quad x \leq h(\alpha, t_L, s) \end{cases} \tag{16}$$

Given this, the analysis in Section 3 can be repeated essentially unchanged to conclude that the two candidate equilibria are $(s = 1,\ x = h(\alpha, t_L, 1))$ and $(s = p,\ x = h(\alpha, t_H, p))$ (we continue to denote the first equilibrium as NC, the second as LC). To confirm that NC is an equilibrium we need to rule out profitable deviations, i.e. we require:

$$h(\alpha, t_L, 1) \geq p h(\alpha, t_H, 1). \tag{17}$$

To confirm that LC is an equilibrium, in a similar way we require:

$$p h(\alpha, t_H, p) \geq h(\alpha, t_L, p). \tag{18}$$

It is easy to prove that, as with our baseline formulation of the cost of cheating, for each $\gamma \in (0, 1)$ there is a non-empty, open set of parameters such that these two inequalities are both satisfied. To do this, let us first note that, if we define

$$g(\alpha, t_L, t_H, p) = \frac{h(\alpha, t_L, p)}{h(\alpha, t_H, p)} \tag{19}$$

we can rewrite the two conditions (17) and (18) as

$$\begin{aligned} p &\leq g(\alpha, t_L, t_H, 1) \\ p &\geq g(\alpha, t_L, t_H, p). \end{aligned} \tag{20}$$

[35] For notational simplicity we neglect to make explicit the dependence of h on γ.

[36] Note that for $s > 0$ equation (15) has a unique solution, which is represented graphically by the intersection between the 45-degree line and a concave function that crosses the vertical axis at $(1-\alpha)s > 0$ and whose value for $x = 1$ is $\alpha t_\tau + (1-\alpha)s$, which is smaller than 1 and larger than $(1-\alpha)s$. The solution is unique, since the function $\alpha t_\tau x^\gamma + (1-\alpha)s$ is strictly concave, starts above and ends below the 45-degree line, therefore it crosses it only once. For $s = 0$ there are two solutions, one for $x = 0$ and one for $x = (\alpha t_\tau)^{\frac{1}{1-\gamma}}$. However, we can rule out the solution at 0 since, by a familiar argument, the offerer would secure a larger expected gain by offering $(\alpha t_L)^{\frac{1}{1-\gamma}}$.

[37] This is obvious from the graphical construction of the solution, and can be verified by computing the derivative of h with respect to t, which is $\frac{\alpha h^{1+\gamma}}{\alpha t h^\gamma (1-\gamma) + (1-\alpha)s} > 0$.

Fix $(\alpha, t_L, t_H) \in (0,1)^3$, with $t_L < t_H$. We have that $0 < g(\alpha, t_L, t_H, 0) = (\frac{t_L}{t_H})^{\frac{1}{1-\gamma}} < 1$. Moreover, $g(\alpha, t_L, t_H, 1) = \frac{h(\alpha, t_L, 1)}{h(\alpha, t_H, 1)} < 1$. The function g, interpreted as a function of p, is continuous and differentiable (as a ratio of continuous and differentiable functions). By direct calculation we see that $g(p)$ is increasing in p.[38]

The function $g(p)$, therefore, is continuous and increasing, its value at 0 is positive (and smaller than 1), its value at 1 is smaller than 1. It therefore crosses the 45-degree line. Let p^* denote the largest value of p such that

$$p^* = g(\alpha, t_L, t_H, p^*). \tag{21}$$

Since g is increasing in p, with $g(0) > 0$ and $g(1) < 1$, for all $p > p^*$ it must be that $g(\alpha, t_L, t_H, p) < p$. Moreover, $p^* = g(\alpha, t_L, t_H, p^*) < g(\alpha, t_L, t_H, 1)$. Therefore, for all $p \in [p^*, g(\alpha, t_L, t_H, 1)]$, both inequalities in (20) are satisfied. Since the closed, non-empty interval $[p^*, g(\alpha, t_L, t_H, 1)]$ contains an open set of values for p, this proves our claim.

Indeed, if the function $g(p)$ were to cross the 45-degree line more than once, there would be also other disjoint non-empty sets of values for p such that $g(p) \leq p$, and given that $g(p)$ is increasing, for all p in those intervals it would remain true that $p \leq g(1)$, so both inequalities in (20) would be satisfied.[39]

It can also be verified that the elasticity of h with respect to s is smaller than 1, and therefore the argument showing that in the multiple equilibria case the correlation between x and s is negative, carries through.

The case $\gamma = 1$ (when the idiosyncratic part of the cost is proportional to the size of the transaction) is a knife-edge: it is still true that the equilibria of our model are one of the three varieties: single *NC* or *LC* equilibrium, and multiple equilibria *NCLC*. The latter, however, can only occur in the non-generic case where

$$1 - \alpha t_H = p(1 - \alpha t_L). \tag{22}$$

This can be easily verified, as the case $\gamma = 1$ can be solved analytically, much as in our baseline specification.

The case $\gamma > 1$, when the (idiosyncratic part of the) cost of cheating increases with the size of the transaction more than proportionately, is qualitatively different. This is intuitive, since if the cost of cheating increases with the size of the transaction in a convex way, there will always be a transaction large enough to generate a cost more than offsetting the gain from cheating, which rises proportionally. The offerer, anticipating this, will always offer a large enough x and the only possible equilibrium will be one without any cheating. To confirm this formally, note that the inequality determining whether or not the $\mathcal{R}$ player of type τ will cheat, i.e.

$$x > \alpha\, t_\tau x^\gamma + (1-\alpha)\, s, \tag{23}$$

now identifies (in principle) two thresholds, increasing with t_τ, with cheating occurring when x is between these two values, and not occurring when x is smaller than the lower threshold (as before) but *also when it is larger than the upper one*. Therefore the $\mathcal{O}$ player will optimally chose an x which is above the upper threshold for the H-type, hence above the upper threshold for the L-type as well, and the only equilibrium is one in which nobody cheats and the transaction is the largest possible.

In fact, the largest threshold, given our normalization that $(\alpha, t_L, t_H, \gamma) \in (0,1)^3$, is above 1, and so it would be ruled out by our assumption than $x \leq 1$. But whereas in the previous analysis this restriction was without loss of generality (since the $\mathcal{O}$ player would never send an offer larger than $\alpha\, t_H + (1-\alpha) < 1$), now it would be rather arbitrary. For the sake of comparison with our earlier results, we nevertheless consider the possible equilibrium outcomes if we restrict x to be smaller than 1 (so that there is only one threshold smaller than 1, for each type τ, below which the $\mathcal{R}$ player does not cheat). In this case the equilibria of the game would be of either the *NC* or the *LC* variety, while the possibility of multiple equilibria would be ruled out. To see this, note that the derivative of g with respect of p is[40] negative when $\gamma > 1$. Hence $g(p^*) > g(1)$, and there is no p such that both inequalities in (20) are satisfied: either $p \leq g(1)$ (supporting the *NC* equilibrium) or $p \geq g(p)$ (supporting the *LC* equilibrium).

5.2. Social cost of cheating depending on the transaction size

The other possibility, mentioned at the beginning of this Section, is to allow the social component of the cost of cheating to depend on x. It can be shown that if we wanted to motivate our cost of cheating as resulting from guilt aversion, in the spirit of Battigalli and Dufwenberg (2009, 2007), we would need to specify the cost function as[41]

$$c(\tau, s, \alpha) = \alpha t_\tau + (1-\alpha) s x \tag{24}$$

[38] The derivative of g with respect to p is $\frac{(1-\alpha)(1-\gamma)h_L(h_H - h_L)}{h_H(h_H(1-\gamma)+\gamma(1-\alpha)p)(h_L(1-\gamma)+\gamma(1-\alpha)p)} > 0$, where for notational convenience h_τ, $\tau \in \{L, H\}$ denotes the function $h(\alpha, t_\tau, p)$.

[39] Although we could not formally prove that the function g has only one crossing, in all the numerical simulations the function turned out strictly concave, with just one crossing.

[40] See footnote 38.

[41] We are grateful to an anonymous referee for pointing this out.

Guilt aversion in our set up is the second order belief of the $\mathcal{R}$ player about the difference between the expected return that the $\mathcal{O}$ player anticipated and the actual return that she obtained because of the action taken by the $\mathcal{R}$ player himself. Since the expected return for the $\mathcal{O}$ player is given by the product between the probability of not being cheated (s) and the offer (x), we obtain a specification in which the dependence on x appears in the social part of the cost.

To pursue this possibility, we explore the more general specification

$$c(\tau, s, \alpha) = \alpha t_\tau + (1-\alpha) s x^\gamma \tag{25}$$

which nests the three possible cases, allowing for the dependence on x to be proportional ($\gamma = 1$), concave ($\gamma < 1$) or convex ($\gamma > 1$), as we did in Section 5.1. The results are broadly unchanged. When $\gamma < 1$ the qualitative results of our baseline specification hold true. In much the same way as before, we can show that the game admits an equilibrium with no cheating (NC), with $s = 1$ and $x = h(\alpha, t_L, 1)$, an equilibrium in which only the L-type cheat (LC) with $s = p$ and $x = h(\alpha, t_H, p)$, as well as a multiple equilibrium regime, with both possibilities, where $h(\alpha, t_H, s)$ is the unique solution of

$$h(\alpha, t_\tau, s) \equiv \alpha t_\tau + (1-\alpha) s h(\alpha, t_\tau, s)^\gamma. \tag{26}$$

Moreover, the set of parameters supporting the $NCLC$ (multiple) equilibrium contains an open subset.

When $\gamma = 1$, again we obtain a knife-hedge case in which multiple equilibria are a non-generic possibility; the restriction on the parameters which must be satisfied to support this possibility is $t_L = p t_H$. Finally, the case $\gamma > 1$ is qualitatively different, in much the same way as in Section 5.1.

5.3. Reciprocity

Yet another possible reason restraining the cheating behavior of the $\mathcal{R}$ player is a concern for (intrinsic) reciprocity, as in Rabin (1993), Dufwenberg and Kirchsteiger (2004), Sobel (2005). According with this view, the $\mathcal{R}$ player would want to reciprocate kindness with kindness, and unkindness with unkindness. Following Dufwenberg and Kirchsteiger (2004), the kindness of the offer x made by the $\mathcal{O}$ player would be measured by the difference between the material outcome[42] she expects the $\mathcal{R}$ player to obtain as a result of her sending x and the simple average between the best and the worst material expected outcomes she could have produced for him (we continue to assume that $x \leq 1$).

In our set-up this boils down to:

$$x(\pi_\mathcal{O} + 2(1-\pi_\mathcal{O})) - \frac{\pi_\mathcal{O} + 2(1-\pi_\mathcal{O})}{2} = (2-\pi_\mathcal{O})(x - \frac{1}{2}), \tag{27}$$

where we denote by $\pi_\mathcal{O}$ the probability that the $\mathcal{O}$ player assigns to the $\mathcal{R}$ player not cheating.

What matters for the behavior of the $\mathcal{R}$ player is his expectation of the $\mathcal{O}$ player's kindness, which depends on second order beliefs $\pi_{\mathcal{R}\mathcal{O}}$, i.e. the expectation by the $\mathcal{R}$ player of the probability that the $\mathcal{O}$ player assigns to $\mathcal{R}$ not cheating. We can then write the $\mathcal{R}$ player expectation of the $\mathcal{O}$ player's kindness as:

$$(2-\pi_{\mathcal{R}\mathcal{O}})(x - \frac{1}{2}). \tag{28}$$

Symmetrically, the kindness of the $\mathcal{R}$ player to $\mathcal{O}$ (conditional on having received x) is measured by the difference between the material payoff he generates for the $\mathcal{O}$ player with his actual action (either cheat, $g = 1$, or don't cheat, $g = 0$) and the simple average between the best payoff for the $\mathcal{O}$ player (that is x, which she obtains if $g = 0$) and the worst (that is 0, which she obtains when $g = 1$):

$$(1-g)(x - \frac{x}{2}) + g(0 - \frac{x}{2}), \tag{29}$$

i.e. $x/2$ if $g = 0$ and $-x/2$ if $g = 1$.

The utility of the $\mathcal{R}$ player is now defined as the sum of his (material) expected payoff and a term reflecting the reciprocity, given by the product of his own kindness with the kindness he expects to receive from the other; given the "product of sign rule", maximizing utility then entails matching kindness with kindness, and unkindness with unkindness[43]:

$$g(2x + \gamma^\tau_\mathcal{R}[(2-\pi_{\mathcal{R}\mathcal{O}})(x - \frac{1}{2})(-\frac{x}{2})]) + (1-g)(x + \gamma^\tau_\mathcal{R}[(2-\pi_{\mathcal{R}\mathcal{O}})(x - \frac{1}{2})(\frac{x}{2})]), \tag{30}$$

where $\gamma^\tau_\mathcal{R}$ is the sensitivity parameter of the $\mathcal{R}$ player of type $\tau \in \{L, H\}$, with $\gamma^H_\mathcal{R} > \gamma^L_\mathcal{R}$.

In equilibrium it must be that $\pi_\mathcal{O} = \pi_{\mathcal{R}\mathcal{O}} = s$, where s is the equilibrium share of transactions not cheated.

Using this we have that the utility of the $\mathcal{R}$ player is:

[42] By material outcome we mean the payoff of the players without considering possible psychological costs or benefits.

[43] For simplicity we assume that only the $\mathcal{R}$ player is sensitive to reciprocity. The conclusion below would not change had we included a concern for reciprocity also in the $\mathcal{O}$ player utility.

$$2x - \gamma_{\mathcal{R}}^{\tau}\left(\frac{(2-s)(2x-1)}{4}x\right) \tag{31}$$

if $g = 1$ and

$$x + \gamma_{\mathcal{R}}^{\tau}\left(\frac{(2-s)(2x-1)}{4}x\right) \tag{32}$$

if $g = 0$.

With this utility function the $\mathcal{R}$ player of type τ does not cheat if and only if

$$x \geq \frac{2 + \gamma_{\mathcal{R}}^{\tau}(2-s)}{2\gamma_{\mathcal{R}}^{\tau}(2-s)}, \tag{33}$$

where $\gamma_{\mathcal{R}}^{H} > \gamma_{\mathcal{R}}^{L}$ implies that $\frac{2+\gamma_{\mathcal{R}}^{H}(2-s)}{2\gamma_{\mathcal{R}}^{H}(2-s)} < \frac{2+\gamma_{\mathcal{R}}^{L}(2-s)}{2\gamma_{\mathcal{R}}^{L}(2-s)}$. Note that in this framework cheating occurs for small offers, while it does not occur for large ones, the opposite of what happens with our formulation. For consistency with the assumption that $x \leq 1$, we assume that $\gamma_{\mathcal{R}}^{\tau} > 2$ for each τ, to guarantee that the threshold above which cheating does not occur is smaller than 1 (for any s). Therefore, any offer $x \geq \frac{2+\gamma_{\mathcal{R}}^{L}(2-s)}{2\gamma_{\mathcal{R}}^{L}(2-s)}$ would not be cheated by any $\mathcal{R}$ player, which implies that the optimal choice by the $\mathcal{O}$ player is $x = 1$. The only equilibrium is then $s = 1$ and $x = 1$.

6. Enforcement

As we mentioned already, the component of the cost of cheating c that is socially determined — via s — can be reinterpreted as arising from an enforcement technology with limited resources for catching and punishing the $\mathcal{R}$ players who cheat. This is quite different from the social norm interpretation we gave before, but the formalisms are surprisingly close in the two cases.

Assume that there is an enforcement agency with resources $k \in [\underline{k}, 1]$.[44] The parameter k pins down the capacity of the enforcement mechanism, in the sense that a mass k of $\mathcal{R}$ players can be checked and fined. Note that, for simplicity, we assume that the enforcement is perfectly targeted; no resources are wasted on $\mathcal{R}$ players that do not cheat.[45] A mass $\min\{k, 1-s\}$ of $\mathcal{R}$ players who cheat is randomly caught and fined. A useful analogy is that of checks for speeding on the highway. Only cars that are actually above the speed limit are randomly stopped, and when they are stopped they are fined. However, the number of cars that can in fact be stopped and fined is limited by the capacity of the police to deploy its patrol cars.

The speeding analogy is also useful to see intuitively how the mechanics of the social feed-back on the cost of cheating work in this case. Given that the police have a fixed capacity for stopping and fining speeding cars, if a large percentage of the cars on the road actually speed it will be impossible for the police to stop all of them. Other things equal, the probability of being caught and fined will be lower when more cars actually speed. Conversely, when very few cars actually speed, the fixed capacity of the police will ensure that many of them will be caught. In fact, once the mass of speeding cars is equal or less than the police capacity, the probability of being caught for those speeding will be one.

It is convenient to normalize the size of the fine to be equal to one. The probability of being caught and fined for an $\mathcal{R}$ player who cheats is then given by

$$z = \min\left\{1, \frac{k}{1-s}\right\} \tag{34}$$

Since the fine is normalized to one, z is also the expected fine, or the expected cost of cheating, coming from the enforcement mechanism.[46]

In line with our simple model above, we assume that z is combined in a linear way with a cost of cheating arising from the $\mathcal{R}$ player's type.[47] We then obtain that the total cost of cheating now is

$$c(L, s, \alpha) = \alpha t_L + (1-\alpha)z \quad \text{and} \quad c(H, s, \alpha) = \alpha t_H + (1-\alpha)z \tag{35}$$

The logic of Section 3 applies to this reinterpretation of the model virtually unchanged. Only NC and LC equilibria are possible. Condition (8) is still necessary and sufficient for an NC equilibrium to exist.

When $s = p$, we must replace (9) with (35). Hence, condition (10) must be replaced by

[44] We take $\underline{k}$ to be a number strictly between 0 and 1. The convenience of assuming that k be bounded away from 0 will be apparent shortly; see footnote 46 below.

[45] Our results easily generalize to the case in which the proportion of enforcement resources that are targeted towards the mass s of $\mathcal{R}$ players that do not cheat is non-increasing in s.

[46] Note that since $k \geq \underline{k} > 0$, we get that $z = 1$ whenever $s = 1$. See footnote 44 above.

[47] As before, this is just the simplest way of proceeding. It also yields immediate comparability with the set up of Subsection 2.1.

$$p\left[\alpha t_H + (1-\alpha)\min\left\{1, \frac{k}{1-p}\right\}\right] \geq \alpha t_L + (1-\alpha)\min\left\{1, \frac{k}{1-p}\right\} \tag{36}$$

which is now necessary and sufficient for an LC equilibrium to exist.

Proposition 1 still holds provided that we replace condition (10) with condition (36).

Note next that if we set $k = p(1-p)$ the new condition (36) coincides with the old condition (10). Hence we can also conclude that, at least for an open interval of values of k around $p(1-p)$, Proposition 2 still holds.

It is then immediate to see that Propositions 3, 4 and 5 still hold in this case.

It is interesting to track what happens to the features of an LC equilibrium as k changes. We begin with the obvious observation that in the model with enforcement in an LC equilibrium the transaction level is equal to

$$\alpha t_H + (1-\alpha)\min\left\{1, \frac{k}{1-p}\right\} \tag{37}$$

since this is the largest offer that will induce the $\mathcal{R}$ players of type H not to cheat. Hence, as we track the LC equilibrium, an increase in k guarantees both an increase in the individual transaction level, and an increase in aggregate transactions not cheated on since the latter is just equal to (37) multiplied by p.[48]

At this point, it would be tempting to use the model to investigate what is the optimal level of resources devoted to enforcement in the case, for instance, we were interested in maximizing the aggregate transactions that are not cheated on. Following standard procedure, the latter would be determined by comparing the marginal cost of increasing k with its marginal benefit, i.e. the marginal increase in the aggregate non-cheated transaction level of the LC equilibrium as k raises.[49] However, given the simplicity of our model, it would be hard to specify a meaningful marginal cost function for k. Moreover, the measure of aggregate activity as we discussed above is open to discussion given our extremely stylized approach to the problem. On both counts, it would be reckless to base detailed policy conclusion on the exercise.

There is however in our view a case where the conclusion that can be drawn is sufficiently strong to deserve special attention. This is what we turn to next. The argument involves again the idea of an equilibrium switch that we discussed above.

Consider a parameter quadruple (α, p, t_L, t_H) such that condition (8), that guarantees the existence of a high trust NC equilibrium, is satisfied as a strict inequality. Note that this is equivalent to $\alpha(pt_H - t_L)/(1-\alpha) < 1-p$. Hence condition (36), that is necessary and sufficient for a low trust LC equilibrium to exist, is satisfied if and only if k is such that

$$k \leq \mathbf{k}(\alpha, p, t_L, t_H) = \frac{\alpha(p\,t_H - t_L)}{1-\alpha}. \tag{38}$$

Now consider the following policy question. We are given a quadruple (α, p, t_L, t_H) such that condition (8) is satisfied strictly. We are also told that society is in the LC equilibrium and that k is equal (or below and very close) to the threshold level $\mathbf{k}(\alpha, p, t_L, t_H)$ given in (38). We are also told that the marginal cost of an increase in k is *finite*. The question then is whether we recommend a *local* policy change — a small increase in k.

The answer must be "yes". The reason is simply that after the increase in k, condition (36) will be violated. It then follows that the *policy will force an equilibrium switch*. The small increase in k will ensure that society switches from the LC equilibrium to the only remaining one — namely the NC equilibrium. Since the marginal cost of k is finite, a "small" increase in k must carry a correspondingly "small" cost. However, by Proposition A.1 there will be a discrete jump up in the aggregate level of activity. Hence the policy change must be worthwhile.

7. Robustness

Our final goal is to investigate whether the high and low trust equilibria have different robustness attributes — different degrees of resilience to small changes in the environment in which they emerge. We do this returning to the original social feedback interpretation of the model, abandoning the enforcement interpretation of Section 6.

We begin with an observation that stems directly from the fact that our model has discrete idiosyncratic cost of cheating types.[50] As we noted above, in any equilibrium the type of player $\mathcal{R}$ with the lowest idiosyncratic cost of cheating that does not cheat (type L in an NC equilibrium, type H in an LC equilibrium) must be *indifferent* between cheating and not cheating. To see this consider a putative equilibrium in which all types that do not cheat strictly prefer not to cheat than to cheat. Then an $\mathcal{O}$ player could unilaterally increase the offer x by a small amount without changing the set of player $\mathcal{R}$ types that cheat, and thus not changing the probability that her offer is cheated on. Since this would clearly increase her expected payoff we conclude that indifference must obtain in any equilibrium for the type with the lowest idiosyncratic

[48] In terms of the economic "well-being" in the sense on Subsection 4.2, there is again a choice to be made as to how we decide to proceed. For simplicity, we proceed with the aggregate transactions that are not cheated on, but this is largely inessential to our claims. Note also that if we take the fine literally, then this component of the cost of cheating would wash out of any welfare calculations as a transfer from some players to the enforcement agency.

[49] Given (37) the marginal increase is $2p(1-\alpha)/(1-p)$ if $k < 1-p$ and zero if $k \geq 1-p$.

[50] As will be apparent, the fact that we have two types as opposed to n, or even countably many, is immaterial as far as this point is concerned. What matters is that the set of types is not a continuum.

cheating cost. Notice that since the set of types is *discrete* we can also conclude that in any equilibrium the type of $\mathcal{R}$ player (if any) that cheats and has the highest idiosyncratic cost of cheating (type L in an LC equilibrium) must strictly prefer to cheat than to not cheat.

These observations about the marginal cheating and non-cheating types directly imply an asymmetry in the response of any equilibrium to an exogenous shock to the trust level s.[51] If the level of s is exogenously shocked *upwards* by a sufficiently small amount, then the set of cheating types will *not change* if x does not change and the $\mathcal{R}$ players best respond to the shock. This is because for a small enough change the marginal cheating type that cheated before the change in s will still strictly prefer to cheat than not to cheat.

If the level of s is instead exogenously shocked *downwards,* and again x does not change and the $\mathcal{R}$ players best respond to the change, the conclusion is quite different. Since the marginal non-cheating type is indifferent between cheating and not cheating before the shock, no matter how small the change in s, after the shock and the consequent decrease in the cost of cheating she will prefer to cheat than not to cheat. Hence her behavior will switch from not cheating to cheating after the shock.

The asymmetry in the response to a shock in s we have described suggests that equilibria with high trust are less robust than equilibria with low trust. In response to an upward shock in s the "realized" new level of s is unchanged, while in response to a downward shock in s the "realized" new level of s is lower.[52] In some sense, downward shocks to s are self-reinforcing, no matter how small, while (sufficiently small) upward shocks are not. It seems reasonable to conclude that, under some reasonable class of dynamics, moving away from a high trust equilibrium towards a low trust equilibrium is "more likely" than the reverse change.

In a previous version of this paper (Anderlini and Terlizzese, 2009) we modeled explicitly another robustness check that points in the same direction as the one we have just sketched out. For reasons of brevity we only summarize the arguments here. Begin with a parameter quadruple that guarantees we are in the NCLC regime, so that both equilibria are possible. Now imagine that a fraction — say q — of the population of players (the same for $\mathcal{R}$ and $\mathcal{O}$ players) myopically believes that they live in a world where the NC equilibrium prevails, while the remaining players believe that the LC equilibrium prevails.

Their beliefs (together with their types) determine the players' expected payoff maximizing offers, cheating costs and hence cheating behavior. We can now imagine a simple dynamics of "myopic belief revision." Players who are cheated "more often" (with higher probability) than they expect according to their myopic beliefs will revise their beliefs in favor of the LC equilibrium and vice versa if they are cheated less often than expected. We can then ask whether the proportion of NC believers will go up or down as time goes by. In the case examined in some detail in Anderlini and Terlizzese (2009), the proportion q of NC believers converges to zero as time goes to infinity. The critical observation to this effect is that the LC believers have in fact a lower cost of cheating than the NC believers. Hence even if we start with a very large proportion of NC believers q, the LC believers, since they will cheat more by virtue of their lower cheating cost, will force a downward belief revision on the part of those players who believe in NC. Hence, in this simple model of myopic belief revision we also conclude that high trust equilibria are in some well defined sense less robust than low trust ones.

8. Summary and conclusions

Given the basic tension between trusting beliefs (and consequent trusting behavior) and the incentives to cheat, we interpreted the level of trust in a society as the belief which resolves that tension. Specifically, we required trust to coincide with the fraction of transactions that are not cheated upon *in equilibrium*, when beliefs are in fact correct.

In keeping with the view that trust is the result of complex social interactions, we introduced a social feedback mechanism. We set up a simple static model in which one side of a transaction can cheat and walk away with the entire surplus, but must suffer a cost of cheating with a socially generated as well as an idiosyncratic component. The socially generated component captures the idea of a social norm, that makes it more costly to cheat when cheating is less common. This is sufficient to generate a rich pattern of possibilities.

There are two main possible reasons why trust levels across societies differ. One is the *multiplicity* of equilibria in spite of *equal fundamental parameters*. The other is a difference in the fundamental parameters themselves. Surprisingly, under some conditions, these two possibilities are empirically partially identifiable in general, and fully identifiable under some further restrictions. The key variable to identification is the equilibrium size of individual transactions, which must be negatively correlated with the equilibrium level of trust when the source of different trust levels are multiple equilibria.

The social feedback component which characterizes our model can be reinterpreted in two different ways. One is as a psychological cost, depending on the second order belief of the receiving player concerning the expectation of the offering player of not being cheated. The less the offering player expects to be cheated, the larger is the cost the receiver (correctly) perceives. The other is as fine imposed by an enforcement agency, who uses limited resources to catch and punish those players who cheat. Because of the limited resources devoted to enforcement, when the fraction of players who cheat is

[51] We are obviously loose in our argument here since we have not defined the response mechanism formally. Hopefully, it will be clear that the details are largely irrelevant provided some basic properties are satisfied.

[52] Again our argument is necessarily somewhat loose here since we have not specified response process. See also footnote 51.

higher, it is less likely that cheating players are caught. This produces a social feed-back mechanism that is the same as the one in the social norm case. Both reinterpretation of the cost have the same implications for the equilibrium of our model, and all our results continue to hold.

We also explore some alternative formulations of the utility of the receiving player, allowing the cost of cheating to depend also on the size of the transaction; these formulation would be implied by guilt aversion or reciprocity. We show that all our qualitative results continue to hold if the effect of the transaction size on the cost of cheating is concave. If it is convex, instead, the nature of the equilibrium is different and, unless we arbitrarily set a small upper bound to the size of the transactions, the only possible equilibrium is one with full trust and maximal transaction size. Indeed, if the cost of cheating increases more than proportionally with the size of the transaction, the fundamental tension between the gains from trusting a transaction partner and the risk of being taken advantage of — which is the focus of this paper — eventually disappears.

Finally, we report briefly on a couple of robustness checks that indicate that high-trust equilibria are more fragile than low-trust ones.

Appendix A

Proof of Proposition 1. To see that the equilibrium set is always not empty, suppose first that

$$t_L \geq p t_H \tag{A.1}$$

In this case, inequality (8) is clearly satisfied and hence an NC equilibrium exists. Next, suppose that (A.1) is violated. Then the middle term of (11) is positive. Since the first term in (11) is positive and strictly less than the third term in (11), one or both of the inequalities in (11) must be satisfied. Since the first inequality in (11) is the same as (10) and the second is the same as (8), we must have that either an LC or an NC or both equilibria exist. Hence the equilibrium set is always not empty.

To see that only the three NCU LCU and LCNC regimes are possible, we only need to argue that there are no equilibria other than the LC and the NC ones.

As we noted in the text, the solution to (4) consistent with Assumption 1 is unique and can only take the values $x = c(L, s, \alpha)$ or $x = c(H, s, \alpha)$. Together with (6), this is sufficient to prove the claim. □

Proof of Proposition 2. We have already noted that the first inequality in (11) is the same as (10) and the second is the same as (8).

Hence, any parameter quadruple such that

$$\alpha(pt_H - t_L) < (1-\alpha)p(1-p) \tag{A.2}$$

must yield the NCU regime. Hence the set of parameter quadruples that yield the NCU regime must contain an open set.

Moreover, any parameter quadruple such that

$$(1-\alpha)(1-p) < \alpha(pt_H - t_L) \tag{A.3}$$

must yield the LCU regime. Hence the set of parameter quadruples that yield the LCU regime must contain an open set.

Lastly, clearly

$$(1-\alpha)p(1-p) < (1-\alpha)(1-p) \tag{A.4}$$

and hence for an open set of parameter quadruples we can be sure that

$$(1-\alpha)p(1-p) < \alpha(pt_H - t_L) < (1-\alpha)(1-p) \tag{A.5}$$

Therefore, we can be sure that the set of parameter quadruples that yield the NCLC regime also contains an open set. □

Proof of Proposition 3. By assumption, the given parameter quadruple supports the NCLC regime. Hence, (11) must be satisfied. The level of individual transactions in the NC equilibrium is $\alpha t_L + 1 - \alpha$, while in the LC equilibrium it is $\alpha t_H + (1-\alpha)\, p$. Hence it suffices to show that (11) implies

$$\alpha t_L + 1 - \alpha < \alpha t_H + (1-\alpha)p \tag{A.6}$$

Consider the first inequality in (11). Divide both sides by p and rearrange to obtain

$$\frac{\alpha t_L}{p} + 1 - \alpha \leq \alpha t_H + (1-\alpha)p \tag{A.7}$$

which, given that $0 < p < 1$ immediately yields (A.6). □

Lemma A.1. *Fix any arbitrary $x \in (0,1)$. Then there exist parameter quadruples that satisfy (8). This ensures that an NC equilibrium exists, and such that the individual transaction level in the associated NC equilibrium (given by $x_{NC} = \alpha t_L + 1 - \alpha$) is in fact equal to the arbitrarily given x.*

Proof. The claim is obvious by inspection of (8). Simply fix α and t_L so as to ensure that $x_{NC} = x$, as required. Then fix an arbitrary $x_H \in (x_L, 1)$, and finally pick p suitably small so that (8) is satisfied. □

Lemma A.2. *Fix any arbitrary $x \in (0,1)$. Then there exists parameter quadruples that satisfy (10). This ensures that an LC equilibrium exists, and such that the individual transaction level in the associated LC equilibrium (given by $x_{LC} = \alpha t_H + (1-\alpha)p$) is in fact equal to the arbitrarily given x.*

Proof. The claim is obvious by inspection of (10). One way to see this is to observe that we can pick α and p arbitrarily close to 1, and ensure that $x_{LC} = x$ as required by choosing the appropriate level of t_H. Picking x_L sufficiently small is then sufficient to ensure that (10) is met. □

Proof of Proposition 4. To avoid ambiguity we will refer to the first second, third and fourth paragraphs of the statement of Proposition 4 as A, B, C, and D respectively.

A is an immediate consequence of Lemma A.2. B and C are immediate consequences of Lemma A.1.

To prove D, proceed as follows. Let an arbitrary parameter quadruple (α, p, t_L, t_H) such that (10) is satisfied and let the individual transaction level in the associated LC equilibrium be $x_{LC} = \alpha t_H + (1-\alpha)p$.

We need to show that we can find two parameter quadruples as follows. A quadruple $(\alpha', p', t'_L, t'_H)$ such that (10) is satisfied, $p' > p$ and $x'_{LC} = \alpha' t'_H + (1-\alpha')p' < x_{LC}$, and finally a quadruple $(\alpha'', p'', t''_L, t''_H)$ such that (10) is satisfied, $p'' < p$ and $x''_{LC} = \alpha'' t''_H + (1-\alpha'')p'' > x_{LC}$.

To construct $(\alpha', p', t'_L, t'_H)$ starting from (α, p, t_L, t_H) we can increase p by a small amount $\varepsilon > 0$, while decreasing t_H by a small amount $2\varepsilon(1-\alpha)/\alpha$, so that overall individual transaction level decreases as required. If we then decrease t_L by the same amount as t_H, it is immediate that the new quadruple $(\alpha', p', t'_L, t'_H)$ must in fact satisfy (10), as required.

To construct $(\alpha'', p'', t''_L, t''_H)$ starting from (α, p, t_L, t_H), we can decrease p by an arbitrarily small amount $\varepsilon > 0$, and set t''_H arbitrarily close to 1. By inspection, for ε sufficiently small and t''_H sufficiently close to 1, the individual transaction level will increase and (10) will be satisfied, as required. □

Proof of Proposition 5. In the low trust equilibrium the individual transaction level is $\alpha t_H + (1-\alpha)\underline{p}$. In the high trust equilibrium the individual transaction level is $\alpha t_H + (1-\alpha)\overline{p}$. The claim is then immediate from the fact that $\overline{p} > \underline{p}$. □

Proposition A.1. *Consider the equilibrium switch of Proposition 3, for a given parameter quadruple giving rise to the NCLC regime. Then, the equilibrium aggregate level of transactions is lower in the associated low trust LC equilibrium than in the associated hight trust NC equilibrium.*

Proof of Proposition A.1. Since the parameter quadruple gives rise to the NCLC regime, (11) must hold. The second inequality in (11), together with the fact that $p < 1$ immediately gives that

$$p[\alpha t_H + (1-\alpha)p] < \alpha t_L + 1 - \alpha \tag{A.8}$$

Recall that $x_{LC} = \alpha t_H + (1-\alpha)p$ and $x_{NC} = \alpha t_L + 1 - \alpha$, and that the aggregate transaction levels in the LC and NC equilibria are $p\,x_{LC}$ and x_{NC} respectively. Hence (A.8) proves the claim. □

Proposition A.2. *There exist parameter quadruples that give rise to the NCLC regime and such that aggregate welfare is larger in the associated NC equilibrium than in the associated LC equilibrium. There are also parameter quadruples that give rise to the NCLC regime and such that aggregate welfare is lower in the associated NC equilibrium than in the associated LC equilibrium.*

Proof of Proposition A.2. Aggregate welfare in the NC equilibrium is given by

$$W_{NC} = 2(\alpha t_L + 1 - \alpha) \tag{A.9}$$

while aggregate welfare in the LC equilibrium is given by

$$W_{LC} = 2[\alpha t_H + (1-\alpha)p] - (1-p)[\alpha t_L + (1-\alpha)p] \tag{A.10}$$

Since the parameter quadruples we are concerned with all give rise to the NCLC regime, (11) must hold.

Notice next that, provided that t_H and t_L are suitably close to each other, (11) is compatible with quadruples that have a value of p arbitrarily close to 1. By inspection of (A.9) and (A.10) in this case we must have that $W_{LC} > W_{NC}$.

It remains to show that for some parameter quadruples that satisfy (11) we can have that $W_{NC} > W_{LC}$. Notice that, again provided that t_H and t_L are suitably close to each other, (11) is compatible with quadruples that have a value of p arbitrarily close to 0. By inspection of (A.9) and (A.10) in this case we must have that $W_{NC} > W_{LC}$. □

References

Anderlini, L., Terlizzese, D., 2009. Equilibrium Trust. EIEF, Working Paper 09/13. http://www.eief.it/files/2012/09/wp-13-equilibrium-trust.pdf.

Arrow, K.J., 1972. Gifts and exchanges. Philos. Public Aff. 1, 343–362.

Attanasi, G., Battigalli, P., Manzoni, E., 2016. Incomplete-information models of guilt aversion in the trust game. Manage. Sci. 62, 648–667.

Banfield, E.C., 1958. The Moral Basis of a Backward Society. The Free Press, New York.

Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Amer. Econ. Rev. 97, 170–176.

Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 144, 1–35.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10, 122–142.

Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., Pancotto, F., 2016. Amoral familism, social capital, or trust? The behavioural foundations of the Italian North–South divide. Econ. J. http://dx.doi.org/10.1111/ecoj.12292.

Blume, L., 2004. Stigma and social control. Cornell University. Mimeo.

Bohnet, I., Herrmann, B., Zeckhauser, R., 2010. Trust and the reference points for trustworthiness in Gulf and Western countries. Quart. J. Econ. 125 (2), 811–828.

Butler, J., Giuliano, P., Guiso, L., 2016. The right amount of trust. J. Eur. Econ. Assoc. 14, 392–436.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74, 1579–1601.

Dixit, A.K., 2003. Trade expansion and contract enforcement. J. Polit. Economy 111, 1293–1317.

Dixit, A.K., 2004. Lawlessness and Economics—Alternative Modes of Governance. Princeton University Press, Princeton, NJ.

Dufwenberg, M., 1996. Time consistent matrimony with endogenous trust. CentER for Economic Research. Mimeo.

Dufwenberg, M., 2002. Marital investments, time consistency and emotions. J. Econ. Behav. Organ. 48, 57–69.

Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., Sobel, J., 2011. Other-regarding preferences in general equilibrium. Rev. Econ. Stud. 78, 613–639.

Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games Econ. Behav. 47, 268–298.

Feddersen, T., Sandroni, A., 2006. The calculus of ethical voting. Int. J. Game Theory 35, 1–25.

Fehr, E., 2009. On the economics and biology of trust. J. Eur. Econ. Assoc. 7, 235–266.

Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does fairness prevent market clearing? An experimental investigation. Quart. J. Econ. 108, 437–459.

Fukuyama, F., 1995. Trust: The Social Virtues and the Creation of Prosperity. The Free Press, New York.

Genakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1, 60–79.

Glaeser, E.L., Laibson, D.I., Scheinkman, J.A., Soutter, C.L., 2000. Measuring trust. Quart. J. Econ. 115 (3), 811–846.

Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development. Amer. Econ. Rev. 94, 526–556.

Horst, U., Scheinkman, J.A., 2006. Equilibria in systems of social interactions. J. Econ. Theory 130, 44–77.

Huang, P.H., Wu, H.-M., 1994. More order without law: a theory of social norms and organizational cultures. J. Law, Econ., Organ. 10, 390–406.

Irlenbusch, B., 2006. Are non-binding contracts really not worth the paper? Managerial Dec. Econ. 27, 21–40.

Kim, J.-I., Lau, L.J., 1994. The sources of economic growth of the East Asian newly industrialized countries. J. Japanese Int. Economies 8, 235–271.

Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. Quart. J. Econ. 112 (4), 1251–1288.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 1997. Trust in large organizations. Amer. Econ. Rev. 87 (2), 333–338.

Mailath, G., Samuelson, L., 2006. Repeated Games and Reputations: Long Run Relationships. Oxford University Press, Oxford.

Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 2, 181–240.

Putnam, R.D., 1993. Making Democarcy Work: Civic Traditions in Modern Italy. Princeton University Press, Princeton, NJ.

Rabin, M., 1993. Incorporating fairness into game theory and economics. Amer. Econ. Rev. 83, 1281–1302.

Sapienza, P., Toldra, A., Zingales, L., 2013. Understanding trust. Econ. J. 123, 1313–1332.

Sobel, J., 2002. Can we trust social capital? J. Econ. Lit. 40, 139–154.

Sobel, J., 2005. Interdependent preferences and reciprocity. J. Econ. Lit. 43, 392–436.

Solow, R.M., 1995. But verify. The New Republic, September 11. Review of Fukuyama's Trust: The Social Virtues and the Creation of Prosperity, pp. 36–35.

Tabellini, G., 2008. The scope of cooperation: values and incentives. Quart. J. Econ. 123, 905–950.

Young, A., 1994. Lessons from the East Asian NICS: a contrarian view. Europ. Econ. Rev. 38, 964–973.