Trade Expansion and Contract Enforcement

by
Avinash Dixit
Princeton University

First draft – March 7, 2002
This version – January 20, 2003

Abstract

I consider a world where traders are separated in geographic, economic, or social space. Gains from honest trade are greater the larger the distance between the pair of traders, but frequencies of meetings, and spread of information about past cheating, both have a local bias. I find that honest trade is self-enforcing only between pairs of sufficiently close neighbors. Global honesty prevails only if the world is sufficiently small. For intuitively reasonable parameters, the extent of self-enforcing honesty decreases when the world expands beyond this size. If external enforcement can be provided at a cost, this is cost-effective only if the world is sufficiently large, and the net payoff from this may or may not be larger than that of a self-governing small community. Worlds of intermediate sizes fare worst, being too large for self-enforcement but too small for external governance.

Address of author:
Avinash Dixit, Department of Economics, Princeton University,
Princeton, NJ 08544–1021, USA.
Phone: 609-258-4013. Fax: 609-258-6419.
E-mail: dixitak@princeton.edu
Web: http://www.princeton.edu/~dixitak/home

*I thank Lars Svensson for useful discussions, and Alessandra Casella, Gene Grossman, Katherine Hafer, Thomas Palfrey, Anne Sartori, Peyton Young, and especially two referees for their insightful comments on the first draft that led to major changes in the model. This revision was carried out during a year as a Visiting Scholar at the Russell Sage Foundation; it is a pleasure to thank them for their support and hospitality. I also thank the National Science Foundation for research support.
1 Introduction

In most trading relationships, one or both parties have opportunities to cheat in an attempt to get some extra private benefit while reducing the total mutual benefit. To ensure honest compliance with the explicit or implicit contract, and thus make the mutual benefit possible, requires methods and institutions to detect and punish cheating. In modern advanced economies, a well-functioning state legal system performs these functions. But in all countries through much of their history, and some countries even now, the legal system is absent, or corrupt, or too slow, to be usable.¹

The alternative to official law that has been studied most thoroughly is self-governance. This works best with direct bilateral repeated interaction, but much of modern economic activity involves dealing with different partners at different times. Self-governance in a group can work on a multilateral basis if members of the group preserve collective memory of any cheating, and all members punish any miscreant by refusing to trade with him. Kandori (1992) and Ellison (1994) develop pioneering theoretical models of this; Ostrom (1990) and Ellickson (1991) are well-known case studies; Greif (1993, 1994) combines modeling and case studies.²

Both modes of analysis suggest that the necessary dissemination of information and credibility of punishment are possible only in small and close groups. Ostrom (1990) found that none of the groups in her studies which were successful in solving collective choice problems using internal systems of information and norms exceeded 15,000 people. The closeness need not be geographic; it could be the result of being engaged in the same kind of economic activity, or of ethnic ties, as discussed by Greif (1993) and Casella and Rauch (2002). The fact that automatic self-governance works only in small and close groups raises several questions. What determines the limits of self-governance? What happens if trading opportunities expand beyond the close self-governing group? When does the official system of

---

¹For example, 25 million cases are pending in Indian courts, and even if no new ones were filed, it would take 324 years to clear the backlog; see Bearak (2000)

²In this paper I consider trading relationships between pairs selected from a large group. This is also the focus of the Kandori, Ellison, and Greif. Ostrom and Ellickson are concerned with other aspects of enforcement, Ostrom with collective action to manage common pool resources, and Ellickson with protection of property rights. But the underlying dilemmas and the problems of instituting self-enforcing methods of information transmission and punishment of cheaters are similar.
governance become more efficient? What happens at the interface between the two systems? In this paper I address these questions from a theoretical perspective.

Greif (1994, 1997) offers especially interesting contrasting studies of two groups with different internal enforcement systems, and correspondingly different sustainable sizes. Maghribi traders, who relied on multilateral group governance, could not expand the scope of their operations. Genoese traders, who used a more bilateral system with formal governance, fared better as trading opportunities expanded. Greif’s theoretical models are system-specific, and we need a model that can encompass both if we are to understand the reasons for their different outcomes. In a different context, Li (1999) offers an explanation based on differences in the costs of the two types of systems. Self-enforcing “relation-based” groups face rising marginal costs – members added at the margin are almost by definition less well-connected, making it harder to communicate information with them and to ensure their participation in any punishments. Formal “rule-based governance” has high fixed costs of setting up the legal system and the information mechanism, but once these have been incurred, the marginal costs of dealing with strangers are low. Therefore relation-based governance can be better for small groups, but rule-based governance will become better for large groups. This makes intuitive sense, but for a deeper understanding one must make more precise the way in which diminishing returns or rising marginal costs set in as a relation-based system grows larger. This is also necessary to understand what happens at intermediate sizes.

That is my aim in this paper. I use a concept of “distance,” which may be geographic or socio-economic, between traders. Traders more distant from each other are less likely to meet and less likely to hear the news of each other’s trading experiences, but they have greater potential gains from trade with each other, making it worthwhile to try to achieve honest trade despite the worsening of their ongoing interaction and communication.

The main results are as follows. There is an upper limit on the size of the trading world in which global self-governance is possible. In worlds larger than this limit, there is an “extent of honesty”, that is, a critical distance such that each trader acts honestly only when dealing with another trader within this distance. Most interestingly, for intuitively plausible values of the parameters, as the world expands, this extent of honesty not merely fails to expand – it actually shrinks. The reason is that in a larger world, the news of your cheating someone at one edge of your neighborhood is less likely to reach others near another edge of your neighborhood.
If external governance can be achieved by installing, at a cost, a global system to monitor cheating, this is worthwhile only if the world is sufficiently large. This is like Li’s idea about relation-based versus rule-based governance, but arises even though I assume the cost of the external governance system to be proportional to the size of the world, so there is no fixed cost. Even when external governance has only constant (not increasing) returns to scale, it can become preferable in a sufficiently large world because of the diminishing returns to self-governance. Worlds of intermediate size fare worst; they are too large for self-governance but too small to make external governance cost-effective. And the net payoff from honest trading under costly external governance in the full large world may or may not exceed what can be achieved by splitting up the large world into smaller communities each of which can sustain self-enforcing trade.

2 The Model

Here I set up the model and define the equilibrium concept.

The trading world

Consider a continuum of traders, uniformly distributed along a circle of circumference $2L$. The mass of traders per unit length of arc is normalized to 1. I will speak of a trader located at a point instead of the density at a point; I hope the slight loss of rigor is more than compensated by the reduction in pedantry. The distance between two traders is measured as the shorter of the two arc lengths between them, clockwise and counterclockwise.

The circle stands for the spectrum of any relevant socio-economic differences, not only or necessarily geographical ones. Thus the differences among traders could be in any one of several dimensions, for example: (D1) resource endowment, including different types of land, labor, physical or human labor, (D2) technology, (D3) geography and climate, and (D4) kinship, ethnicity, language, culture or religion. The circle is of course a specific way to model this, but the general intuition for the results that emerge from the model is robust, and I will comment on this when discussing the main results.
Matches

There are two periods. The first period is the one where honesty or cheating are the crucial issue; the appropriate rewards or punishment come in the second period, which may as usual stand for the reduced form of a longer future. The payoffs are expressed in present values so no further discounting is necessary. In each period, traders are randomly matched in pairs. The matches are assumed to have the following properties:

(M1) Independence: The actual match in period 1 does not affect the probabilities of matches in period 2.

Comment: This is assumed for algebraic simplicity. In reality, traders may try to build reputations and persevere in matches that have had good outcomes, but so long as there is some exogenous probability of separation due to death or retirement of one of the parties, the qualitative results will be unaffected.

(M2) Localization of matches: In each period, each trader meets exactly one partner. For each trader, the probability of meeting another located at distance \( x \) is

\[
\frac{e^{-\alpha x}}{2 \left[ 1 - e^{-\alpha L} \right] / \alpha}.
\]

The denominator is just a normalizing factor to ensure that the probabilities of matches at all distances between 0 and \( L \) on either side of any one trader sum to 1.

Comment: The parameter \( \alpha \), assumed to be positive, captures the matching technology; the smaller is \( \alpha \), the less localized the technology, and as \( \alpha \to 0 \), the distribution of matches tends to uniformity over the whole circle. Localization is a realistic assumption, even in the internet age. Any one trader may post information about himself on the web, but a match requires someone to check this site. A search engine will typically find thousands of potential sites, and people, being constrained by time, will select the ones they have some familiarity with, namely ones local to them. However, the assumption has a different restrictive aspect, namely that I am leaving such a process of search in the background and specifying the probabilities exogenously in a reduced form. This, and the negative exponential function, are needed for tractability, but it would be an interesting research challenge for the future to endogenize the search and matching process.

(M3) Desirability of trade expansion: The payoffs from a match at distance \( x \) are proportional to \( e^{\theta x} \). (The complete payoff matrix will be specified shortly.)
Comment: The parameter $\theta$, assumed to be positive, captures the benefit from expanding the scope of honest trade. It can be thought of as reflecting the idea of comparative advantage in trade theory. Consider some examples related to the distance concepts (D1)-(D4) listed above. (D1) If the space is socio-economic, with traders located according to the ratio of land to labor they possess, then a landowner will benefit most from meeting someone who primarily owns the complementary input, labor, than from meeting another landowner. (D2) An electrical engineer will stand to gain more from meeting another engineer with a complementary technology such as mechanical or chemical than from meeting another electrician. (D3) Geographical and climatic differences enable people to grow different crops and trade to benefit from variety in consumption. In some other respects, for example local public goods or goods for which tastes are culture-specific, interaction may be more beneficial the more similar the partners are; this is like item (D4) in that list. Such trades will unambiguously best be carried out using automatic self-governance in small communities of similars. But governance of global trade is regarded as an important concern in the real world; therefore there must be significant transactions whose potential benefits rise with expansion of the scope of trade. My focus is on those transactions. Again, the exponential form is special, but the qualitative results are unaffected by this choice.

(M4) Localization of information: If a trader in a match cheats the other, the probability that a third person, located at distance $y$ from the victim of the cheating, receives news of this cheating is $e^{-\beta y}$.

Comments: [1] The parameter $\beta$, assumed to be positive, is a feature of the communication technology. The choice of negative exponentials is special, but natural. If one thinks of contact and communication decaying at a constant rate with distance, then a geometric or negative exponential form will result. [2] I am assuming that news travels from the victim to any third person along the shorter arc of the circle connecting them. If news travels in both directions, then the probability of it reaching someone distant $y$ from the victim is

$$1 - \left( 1 - e^{-\beta y} \right) \left( 1 - e^{-\beta (2L-y)} \right).$$

For any fixed $y$, as $L$ goes to infinity, this goes to $e^{-\beta y}$, the same as my unidirectional probability. Therefore my results for large $L$ are entirely unaffected by my assumption. For smaller $L$, two-way flow would increase the probability of spread of news and therefore reduce the incentive to cheat, but the qualitative results remain unaffected.
Figure 1: Locations, matches, and information transmission

Figure 1 illustrates the various concepts and magnitudes.

What are the plausible relative magnitudes of the three parameters $\alpha$, $\beta$ and $\theta$? First, $\theta > 0$ is important for the existence of gains from expanding the scope of trade. Next, we need $\alpha > \theta$ for convergence of expected values when $L$ is large. To simplify some later notation, I define $\lambda = \alpha - \theta > 0$. Finally, if the technologies of transportation and communication are intrinsically linked, then $\alpha$ and $\beta$ will have similar values and will shift together. At one time, meeting and communication were literally the same thing. Even in the internet age, face-to-face meeting and communication remain important, and the World Wide Web itself is a common medium for matching and communication activities, so we should expect $\beta$ and $\alpha$ to have similar values.\(^3\) This is not maintained as an assumption; I will use it to guide our thinking about alternative logically possible cases of results. In specific contexts, econometric analyses similar to those of Conley and Udry (2001) on technology transmission could be used to estimate these parameters.

We have a degree of freedom to choose the unit in which distance is measured. In fact what matters is not physical (or socio-economic) distance as such, but the probabilities $e^{-\alpha L}$ and $e^{-\beta L}$ and the size of the gain $e^{\theta L}$. If the unit of distance is halved, $L$ is doubled but the parameters $\alpha$, $\beta$ and $\theta$ are all halved, leaving $\alpha L$, $\beta L$ and $\theta L$ unchanged. An implication for comparative statics is that all the crucial magnitudes, for example the maximum size of the world compatible with global honest self-governance, are homogeneous of degree $-1$ jointly in the three parameters $\alpha$, $\beta$ and $\theta$.

\(^3\)See The Economist (2002) for a discussion of the continuing importance of physical meeting and communication in the internet age.
Player types, information, strategies, payoffs

The traders come in two behavioral types of players, called Normal or N-type and Machiavellian or M-type; the latter should be thought of as especially skillful cheaters. Both types uniformly distributed along circle. Their proportions in the population are \((1 - \epsilon)\) for the N-types and \(\epsilon\) for the M-types. The number \(\epsilon\) is assumed to be small. The type of each player is not directly observable to others, but they may condition the probabilities of a person’s type based on any observables. It will be seen that the role of the few Machiavellian traders is to pin down expectations out of equilibrium and to prevent a cheater from attempting to escape punishment by claiming that the cheating was just an error.

In each period, IID random matchings are made, as specified in (M1) above, to determine trading pairs. Each player in a pair knows the other’s distance \(x\), and with the information mechanism specified in (M4), may know the other’s past history of cheating if any. The stage game of each matched pair proceeds as follows:

[G1] Each decides whether to play. These choices are simultaneous. The outside opportunity for each is normalized to 0. If both choose to play, the outcomes depend on their actions and types as specified next.

[G2-n] When two N-types meet – this accounts for most of the matches – in period \(t\), the game has a payoff matrix \(\exp[\theta x]\) times

<table>
<thead>
<tr>
<th></th>
<th>Trader N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honest</td>
<td>(h_t, h_t)</td>
</tr>
<tr>
<td>Cheat</td>
<td>(w_t, s_t)</td>
</tr>
</tbody>
</table>

\(\theta\)

[G2-m] If an N-type meets an M-type, the N-type gets \(s_t\) regardless of his own action, and the M-type gets a positive payoff. If two M-types meet, they each get positive payoffs.

The payoffs are assumed to satisfy the following inequalities:

(P1) \(w_t > h_t > c_t > s_t\). This simply says that the stage game between two N-types is a prisoners’ dilemma.

(P2) \((1 - \epsilon) c_t + \epsilon s_t > 0 > s_t\). This says that an N-type will choose to play, rather than take the outside opportunity, against random opponent even if the latter is going to cheat, but not against a known M-type. The two together create an expected cost of cheating for an N-type: if you cheat in period 1, and your period-2 match gets to know of this, he will
not play you, and you will get 0 instead of $c_2$. This device for achieving cooperation with finite repetitions is similar to that in Benoit and Krishna (1985). Note that the assumption places an upper bound on $\epsilon$. (If $\epsilon$ is very small, (P2) is approximately $c_t > 0 > s_t$.)

(P3) $w_1 - h_1 < c_2$. This says that in a distanceless world, if cheating is detected and publicized with certainty, then there exists an equilibrium where all N-types choose Honest in period 1. (To be more precise, both sides of the inequality should be multiplied by $(1 - \epsilon)$.)

The role of this assumption is to confine the question of honest behavior to the context of uncertain dissemination of information across distance that is the focus here. To simplify later algebra, define $k = (w_1 - h_1)/c_2$, then (P3) is equivalent to $k < 1$.

**Equilibrium**

The solution concept is Perfect Bayesian Equilibrium. This can be characterized by a number $X$, such that Normal types behave honestly in period 1 when meeting someone located within distance $X$ of themselves, and cheating otherwise. To be more precise, the equilibrium strategy for N-types is as follow:

In Period 1 choose to play, and choose the action Honest if the partner’s location is $X$ or less away from you, and Cheat if between $X$ and $L$.

In Period 2, if you have received information that your current match’s period-1 match received $s_1$, then choose not to play (take the outside opportunity); else play and choose Cheat.

M-type always choose to play, and their strategies can be kept in the background of the whole analysis.

Think of this for now as specifying a “candidate” equilibrium and put it to the test. The only actions and beliefs at issue are those of the N-types. Most aspects of these are obvious. [1] In period 1, when the partner’s type is not known, choosing to play has positive expected payoff by (P2). [2] In period 1, if the partner is located at distance between $X$ and $L$, his strategy specifies that he should play Cheat. Then it is optimal for you to play Cheat also: it gives a better payoff at that time than Honest, and since the partner is not going to receive $s_1$, has no adverse consequences for you in period 2. [3] In period 2, if you have no information about the partner’s cheating in period 1, then choosing to play and cheating are clearly optimal. [4] In period 2, if you have received information that your current partner’s
period-1 match got payoff $s_1$, then given the strategies in the candidate equilibrium, your Bayesian inference is that your current partner is an M-type. Therefore choosing not to play is optimal by (P2). (This is the crucial role of the M-types in the model. Without them, there would be no cheating in equilibrium, and therefore the inferences drawn after a deviation could be arbitrary. A deviator could claim that it was just a mistake and ask the period 2 partner to play for sake of the positive payoff $c_2$. Now the partner is deterred from acceding to such a request by his Bayesian calculation that he will instead get the negative $s_2$.)

It only remains to check whether, or more precisely, under what conditions, it is optimal to act Honest in period 1 if the current match is located at distance $X$ or less. This is done in the following Lemma and Proposition. The formal proofs of these, and of other propositions, are in the Appendix; in the text I state the ideas behind the proof and the intuition.

**Lemma:** Suppose an N-type finds that his period-1 partner is distant $X$ from him. If his belief is that this partner will play Cheat even if an N-type, then his best response is also to Cheat. If his belief is that this partner, if an N-type, will play Honest, then his best response is also Honest if $F(X, L) \geq 0$, and Cheat if $F(X, L) < 0$, where the function $F(X, L)$ is defined by

$$F(X, L) = \frac{\beta e^{-\lambda X} - \lambda e^{-\beta X}}{\beta^2 - \lambda^2} + e^{-(\beta + \lambda)L} \frac{(\lambda e^{\beta X} - \beta e^{\lambda X})}{\beta^2 - \lambda^2} - k e^{\theta X} \frac{1 - e^{-\alpha L}}{\alpha}.$$ (2)

The proof consists of a simple but tedious calculation of the expected future costs of cheating.

**Proposition 1:** For each $L$, there exists a unique $X(L)$ such that an equilibrium with strategies as defined above exists for any $X$ satisfying $0 \leq X \leq X(L)$.

The intuition is that cheating becomes more attractive the more distant the partner. This follows from the localization of matches (M2) and information (M4). If you cheat someone farther away from you, the news is less likely to reach potential future partners closer to you, and they are the ones you are more likely to meet in the future. The assumption that the payoffs increase with distance (M3) works in the same direction but the reasoning is a little more subtle. Cheating someone farther away gets you a larger immediate gain. But this does not affect the distribution of distances of the future partners (M1), and therefore does not affect the cost of cheating. The overall result is that the net benefit from cheating increases.
with the distance of the current partner. Therefore the structure of equilibria, where people behave honestly with others within a certain distance of themselves, is quite intuitive.

The formal proof first verifies that $F$ is a decreasing function of $X$ holding $L$ fixed. Then three cases arise: [1] If $L$ is such that $F(0, L) < 0$, then $F(X, L) < 0$ for all $X$ and it is always optimal for an N-type to cheat in period 1, so the candidate equilibrium is not an actual equilibrium for any $X$, that is, $X(L) = 0$. [2] If $L$ is such that $F(L, L) > 0$, then the candidate equilibrium has self-fulfilling expectations of honest behavior for any $X$ between 0 and $L$, so $X(L) = L$. [3] If $F(0, L) > 0 > F(L, L)$, then by monotonicity there is a unique $X(L)$ in $(0, L)$ such that $F(X, L) \geq 0$ for $0 \leq X \leq X(L)$ and $F(X, L) < 0$ for $X(L) < X \leq L$. The candidate equilibrium is sustained by self-fulfilling expectations in the former range, but not in the latter range because there each trader finds it better to play Cheat even if he expects others to play Honest.

As usual in such games, there is a multiplicity of equilibria, each sustained by its own expectations. But I will give this system its best shot by looking at the best possible $X$ for each $L$, that is, the function $X(L)$ itself. I will call $X(L)$ the extent of honesty in equilibrium. The next three sections examine its properties.

## 3 Gain from Expanding the Scope of Honest Trade

The equilibrium established above brings social benefits because honest trade is sustained in period-1 matches between two N-type people who are located within distance $X$ of each other, resulting in a payoff of $h_1$ rather than $c_1$ from each such match. Leaving aside the constant multiplicative factor $(1 - e)^2 (h_1 - c_1)$, the size of this gain is given by the function

$$V(X, L) = \frac{\alpha}{2(1 - e^{-\alpha L})} 2 \int_0^X e^{-\alpha z} e^{\theta z} dz$$

$$= \frac{\alpha}{\alpha - \theta} \frac{1 - e^{-(\alpha - \theta) X}}{1 - e^{-\alpha L}}$$

This is increasing in $X$ for any fixed $L$, and decreasing in $L$ for any fixed $X$.

In particular, the benefit from sustaining honest trading over the whole world in period-1 is given by

$$V(L, L) = \frac{\alpha}{\alpha - \theta} \frac{1 - e^{-(\alpha - \theta) L}}{1 - e^{-\alpha L}}$$

11
Figure 2 shows this function. Note that $\theta > 0$ is important for it to be increasing, and $\alpha > \theta$ for convergence as $L$ increases.

![Graph](image_url)

Figure 2: Benefits of sustaining honesty over larger circles

### 4 The Limits of Honest Trade

In the discussion following Proposition 1, the logical possibilities for the extent of trade $X(L)$ were seen to range from 0 all the way to $L$. How do the actual outcomes vary with $L$? The localization of matches and information suggests that honesty over the full circle ($X(L) = L$) should be possible for small $L$ but not for large $L$. This is seen more formally from the following:

**Proposition 2:** There exists a unique positive $L^*$ such that $X(L) = L$ for $0 \leq L \leq L^*$, and $X(L) < L$ for $L > L^*$.

The proof consists of examining for what values of $L$ we have $F(L, L) \geq 0$.

Next comes what is perhaps the most interesting finding from this model: for plausible parameter values, as $L$ increases beyond $L^*$, the extent of honesty $X(L)$ decreases. This is the result of numerical calculation backed up by an intuitive explanation, rather than analytical proof, so it may not deserve to be called a proposition, but for clarity of exposition I have labeled it as such.

**Proposition 3:** If $\beta \geq \theta$, then as $X(L)$ is a decreasing function for $L > L^*$.

The intuition for this is seen from Figure 3. The circle on the left is of critical size $L^*$. Therefore a trader located at O, when meeting another at P, is indifferent between honesty and cheating. The circle on the right is somewhat larger: $OP_1 = OP_2 = L^*$, with added
people between. Now, if O meets P₁ and cheats, the probability that P₂ finds out is now < 1; whereas it was = 1 before. So the cost of cheating has decreased. The larger is β, the bigger this effect, because then the news transmission probability decays faster. But if O cheats P₁, he risks that any period-2 meetings he may get with traders located between P₁ and P₂ may become unproductive if they have heard of his cheating P₁. This effect is bigger the larger is θ (because then the potential gains from these meetings are bigger) and/or the smaller is β (because then the probability of news transmission to these traders is higher).

The balance of the two effects depends on β and θ. If β is large relative to θ, the first effect will be stronger and X(L) will decrease as L increases beyond L*. ⁴

Figure 4 shows the (θ, β) combinations where the two effects are exactly equally strong

---

⁴In an earlier version of this paper, matches beyond the extent of honesty were ruled out. Then the second effect was absent, and X(L) decreased as L increased beyond L* for all parameter values. But the mechanism employed there for ruling out the matches beyond X(L) was defective.
and opposite. The function $X(L)$ is decreasing for all parameter values above this curve, and increasing for all values below it. In particular, since the curve lies entirely below the 45-degree line, $\beta \geq \theta$ is sufficient to ensure a decreasing $X(L)$. Since $\alpha > \theta$, and since $\beta$ and $\alpha$ are likely to be of similar magnitudes as was explained in the comments following the statement of match conditions (M1)-(M4), the plausible case is indeed $\beta > \theta$. Thus $X(L)$ decreasing is the plausible outcome.

The figure has $\alpha = 0.05$ and $k = 0.5$. The former is without loss of generality, because by a suitable choice of the unit for distance we can fix one of the three parameters $\alpha$, $\beta$, and $\theta$ arbitrarily. Calculations for the whole range of $k$ from 0 to 1 yield similar curves. Thus the numerical analysis is exhaustive, but I have not been able to obtain an analytic proof of the finding.

Finally, the extent of honesty in a very large world is found from the following:

**Proposition 4:** As $L \to \infty$, $X(L)$ tends to a positive limit $X^*$ if $k(\beta + \lambda) < \alpha$. Else $X(L)$ reaches zero for a finite $L$.

To sum up, Figure 5 shows a typical $X(L)$, taking its decreasing beyond $L^*$ as the plausible case, and allowing it to asymptote to a positive level.

![Figure 5: Best equilibrium $X$ as function of $L$](image)

The decline of honesty as the size of the world grows beyond the limit of global self-governance was explained intuitively using Figure 3. This makes it appear crucially dependent on the assumption that the traders are located along a circle. However, the general idea behind the picture is much more robust. All we need is that people have a limited capacity for contacting others to convey or to receive information. (Even in the internet age, one can check only a few of all available sites that mention “cheating”.) Take any other structure of
spatial location, for example a disc in two dimensions or a sphere in three dimensions, and
consider a trader, call him A, located at a boundary of this. All of his available contacts are
somewhere inside the disc or the sphere, and are starting links in a chain of communication
that leads to another trader B. Now let the space expand, so A becomes interior to the new
bigger disc or sphere, and his contacts get redistributed. Some of the older contacts will
become weaker or wither away, and be replaced by new contacts with new neighbors in the
added regions of the space. Some of these new contacts may become part of a new chain that
reenters the old part of the space and leads to the same old B. But these new paths travel
through the added outer regions of the space. They must therefore be on average longer
than the old ones, and therefore information will travel less well from A to B than it did
before. In other words, any two traders in the original trading world will become less well
connected as the world expands. This will lead to the same kinds of results as the circle.

5 Comparative Statics

Here I examine the dependence of $L^*$, the largest world size compatible with self-governance,
and $X^*$, the extent of honesty in a very large world, on the parameters $\alpha$, $\beta$, $\theta$, and $k$. First,
from the freedom to choose the unit in which distance is measured, it is obvious that

**Proposition 5:** Each of $L^*$ and $X^*$ is homogeneous of degree $-1$ jointly in $(\alpha, \beta, \theta)$ for
any given $k$.

Next, turning to the effect varying the parameters one at a time, we have

**Proposition 6:** $L^*$ is a decreasing function of each of $\alpha$, $\beta$, $\theta$, and $k$.

The intuition is that greater localization of matches or information, and greater benefit
from cheating a partner far away, all make cheating more attractive and make it harder to
sustain self-enforcing honesty. Likewise, an increase in the ratio of the immediate gains from
cheating to the value of period-2 trade, $k = (w_1 - h_1)/c_2$, increases the incentive to cheat.

**Proposition 7:** $X^*$ is a decreasing function of each of $\beta$ and $k$.

The intuition is the same as for $L^*$. Similar results cannot be obtained for $\alpha$ and $\theta$
because they affect the probabilities and values of matches beyond $X^*$ in different ways.

The homogeneity of degree $-1$ result suggests a useful way of displaying numerical results.
We can regard the equation which defines $L^*$ as linking three magnitudes $p$, $q$, and $h$ defined
as follows:
(i) \( p \equiv e^{-\alpha L^*} \) is the probability of meeting a trader at the far end of the world relative to that of meeting an immediate neighbor,

(ii) \( q \equiv e^{-\beta L^*} \) is the probability that an act of cheating a trader at the far end of the world becomes known to one’s immediate neighbor,

(iii) \( h \equiv e^{\theta L^*} \) is the potential gain from trading with someone at the far end of the world relative to that of trading with an immediate neighbor.

I examine the relationships between these magnitudes pairwise. First, consider \( p \) and \( q \) holding \( h \) fixed. In fact I hold \( \theta = 0 \) so \( h = 1 \); this enables me to focus on the variation in probabilities with distance, leaving aside the issue of how the size of gains varies with distance. I choose \( k = 0.5 \).

![Figure 6: Probabilities in the maximal self-governing group](image)

Figure 6 shows the relation between \( q \) and \( p \) at the limit of full self-governance. The curve is the lower boundary of self-governance; any point to the right or above it allows global honesty. We see a tradeoff between \( p \) and \( q \) along the boundary: if a trader is relatively more likely to meet others far distant, then a smaller probability of dissemination of the news of cheating over that distance suffices to preserve honesty. However, we see that \( q \) is the relatively more important probability; its variation is confined to a tight range as \( p \) varies over its full range. In other words, the rate of communication decay \( \beta \) is more important than the localization of meetings captured by \( \alpha \).

We also see that the range of \( q \) is somewhat under 0.5. With \( \theta = 0 \), \( k \) is just the ratio of immediate gains from cheating to the future cost if discovered. Therefore in a locationless
model, any probability of discovery exceeding $k$ would be enough to dissuade cheating. Here $k = 0.5$, but a somewhat smaller probability of the spread of news across the maximal distance in the community is needed. The average of the probabilities over all traders in the circle is of course larger.

A similar calculation can be carried out for the probabilities $P \equiv e^{-\alpha X^*}$ and $Q \equiv e^{-\beta X^*}$, for the extent of honesty in a large world. The result is shown in Figure 7. Here $Q$ and therefore the news parameter $\beta$ is relatively more important when $P$ is small or when $\alpha$ is large, but when $\alpha$ is small, the two parameters are of comparable importance.

![Figure 7: Probabilities for the extent of honesty in a large world](image)

Next consider the tradeoff between $q$ and $h$ (or equivalently, between $\beta$ and $\theta$), holding $p$ constant. The equation that determines $L^*$ (in the proof of Proposition 2 in the appendix) can be rewritten as

$$\frac{1 - p/(qh)}{\ln(p) - \ln(qh)} - \gamma \left( \frac{1/p}{\ln(p)} - \frac{1}{\ln(p)} \right) = 0.$$ 

Therefore the relation between $q$ and $h$ is simply $qh = \text{constant}$. Similarly for $Q$ and $H$. In this sense, the parameters $\beta$ and $\theta$ are equally important.

## 6 External Enforcement

In this section I compare the self-enforcing governance analyzed so far with a formal or official method. Suppose that at a cost $c$ per unit of arc length along the circle, any cheating can be detected and the information made available to future traders. This should be thought
of as an auditing or monitoring system, which future traders can consult to find out their potential partners’ histories. This could be a credit-history agency or a trade association set up by the group of traders, or a state legal system. To give this alternative its best chance (just as I gave the self-governance system its best chance by focusing on the maximal possible $X$, namely $X(L)$), I will assume that it is implemented by a benevolently to maximize the representative trader’s payoff. I will assume that the detection system covers the whole circle, and that its costs are recovered from the traders by levying a lump sum charge $c$ on each of them. (Remember that the mass of traders per unit arc length has been normalized to 1.) Under this system, when the semi-circumference of the circle is $L$, the payoff for each trader will be $V(L, L) - c$, where the function $V$ is defined by (4).

How does such a system of external enforcement compare with the self-enforcement system studied in the previous sections? To make the comparison, let us first find the payoff under self-enforcement as a function of $L$. When $L \leq L^*$, self-governance is possible over the full circle, yielding payoff $V(L)$ to each trader. When $L$ increases beyond $L^*$, in our plausible case the maximum feasible extent of honesty $X(L)$ decreases. Then the payoff $V(X(L), L)$ decreases since $V(X, L)$ is increasing in its first argument and decreasing in its second argument. Eventually the payoff falls to $V(X^*, \infty)$.

Figure 8: Optimal enforcement modes in different-size worlds

Figure 8 compares the two payoff functions. The figure shows the gross and net payoffs from external enforcement; these are the two parallel curves $V(L, L)$ and $V(L, L) - c$. It also shows a falling curve for self-enforcement beyond $L^*$, starting at $(L^*, V(L^*, L^*))$ and going
to $(\infty, V(X^*, \infty))$. The thick curve in three separate segments is the payoff function that arises from choosing the better of the modes of enforcement for each $L$.

When $L \leq L^*$, self-enforcement is globally effective, and saves the detection cost $c$, so it is obviously superior to external governance. Beyond $L^*$, there is an interval where the payoff from self-enforcement falls below the gross payoff $V(L, L)$ from external enforcement but remains above the net payoff $V(L, L) - c$ of that system. Thus self-enforcement no longer works over the whole circle, but external enforcement is not yet cost-effective. So long as

$$c < V(\infty, \infty) - V(X^*, \infty) = \frac{\alpha}{\alpha - \theta} - V(X^*, \infty).$$

the rising curve $V(L, L) - c$ for external enforcement and the falling curve $V(X(L), L)$ for self-enforcement eventually cross. Beyond that point, external governance is preferable. However, its eventual payoff $V(\infty, \infty) - c$ may or may not climb back up to $V(L^*, L^*)$, the best possible payoff under self-governance. It will do so if

$$c < V(\infty, \infty) - V(L^*, L^*) = \frac{\alpha}{\alpha - \theta} - V(L^*, L^*).$$

If this is not the case, then it is preferable (if it is feasible) to split up a large world into smaller self-contained circles of size $L^*$ each.

In other words, small communities can achieve full self-governance using their own information systems and do not need external governance. In very large communities, the benefits that are available from trade with distant partners can only be realized by instituting a system of external governance at a cost. Communities of an intermediate size fare worst: they are too large for self-governance but too small for external governance. When an expanding economy reaches the size where external governance becomes just cost-effective, “it is darkest just before the dawn” for it. However, even very large communities with external governance may or may not be better than the optimal-sized self-governing small communities.\(^5\)

The idea that small-sized self-governing groups may be better than large anonymous markets has been developed in recent literature in other contexts, which thus complement the analysis of this paper. Kranton (1996) constructs a model in which individuals can choose

---

\(^5\)To put this more precisely, the function composed of the thicker segments in Figure 8 has two local maxima, at $L^*$ and $\infty$, either of which may be the global maximum depending on the parameters. The function has a local minimum at the point where the payoff curves for self-governance and external enforcement cross; its global minimum is at 0.
between bilateral repeated interactions, which must be self-enforcing, and an anonymous market, which requires search but does not have enforcement problems. The market is the outside opportunity that defines the incentive compatibility constraint for the former. If more people choose the market, it is thicker and offers better prospects for search. This tightens the incentive-compatibility constraint for bilateral relationships and therefore makes it harder to sustain self-enforcing cooperation in them. Conversely, if more people engage in reciprocal bilateral exchange, the market is thinner, bilateral exchange can be sustained more easily, and offers higher payoff. Either system can prevail in equilibrium even when it is socially less efficient. The model of Bowles and Gintis (2000) is closer to my paper. They consider disjoint networks which are like islands in a sea of an anonymous market. The members of a network trade only among themselves. The larger is the network, the higher is the probability of finding a trading partner, but the worse is the signal about his past behavior. They find the equilibrium and optimal size of a network. They have a richer specification of search and information flow within a network. They allow no information flows between a network and the market, yet migration can take place into and out of a network. By contrast, I do not have disjoint networks but overlapping neighborhoods of honesty for each trader. Both trade and information can go across these neighborhoods. I can also examine the effect of changing the size of the world as a whole. Fafchamps (2002) considers traders each of whom is connected to a subset of the others by a network of contacts, which allows information about the history of a trader to disseminate to others over time. His focus is on the dynamic process by which spontaneous order may emerge, not on comparing difference systems of governance.

7 Concluding Comments

Economic theory traditionally took for granted the existence of a well-functioning official legal system to enforce property rights and contracts. In response to persistent findings to the contrary from many countries and periods of history, modeling of alternative systems is beginning to emerge. This paper contributes to that literature, focusing on self-governing groups whose size is limited by their information flows and punishment mechanisms. Governance can also be provided for a profit by private third parties; recent models of this include Anderson and Bandiera (2000) for property right protection, and Dixit (2003) for
contracts. Considerations of the interaction among the various alternatives and an imperfect formal system of law includes such questions as: When are alternative systems substitutes and when complements? Will evolution and adaptation to changing conditions lead to the emergence of efficient systems of governance, or can there be lock-in that allows inefficient systems to persist? This promises to be a rich set of topics for future research.

Several generalizations of the model seem worth pursuing in the course of such research. Here are some examples: [1] The matching process can be endogenized. Finding a partner in any one period may be itself be a random event with probability less than 1, and the probability may also be a function of distance. [2] The information transmission process can be modeled explicitly and the probabilities of spread of the news of cheating can be endogenized. [3] Even if the matching and information technologies are kept exogenous, different functional forms for them can be investigated. [4] The spread of news of cheating may depend not only on $y$, the distance between the victim and third parties, but also on $x$, the distance between the cheater and the victim, perhaps because victims may feel that their immediate neighbors are unlikely to meet a cheater located far away. [5] The mechanism of external enforcement in Section 6 can be modeled more explicitly. As a simple example, the auditor may be thought of as someone located at the center of the circle who gets news from everyone; then the parameters $\beta$ and $c$ may be related to each other. Other possibilities such as increasing or eventually diminishing returns to scale to external enforcement can also be tried. I hope that the model of this paper will provide a starting point for such extensions, and for more major modifications to address other questions to do with contract enforcement.
Appendix

Formal proofs of various propositions from the text are collected here.

Proof of Lemma:

Focus on any one trader, called Oscar for identification, and located at the point O on the circle shown in Figure 9. If Oscar believes that his period-1 partner will play Cheat if an N-type, then Oscar expects a better period 1 payoff by responding Cheat than by responding Honest. Also, by (P1) and (P2), even if Oscar plays Cheat, Oscar’s partner, whether an N-type playing Cheat or an M-type, is not going to get the bad payoff $s_1$. So the action has no period-2 consequences for Oscar. Thus Cheat is the optimal response for Oscar in this situation.

If Oscar believes that his period-1 partner, located at distance $X$ from him, will play Honest if an M-type, then Oscar stands to gain an immediate

$$(1 - \epsilon) (w_1 - h_1) e^{\theta X}$$

by choosing Cheat instead of Honest. The expected cost of this to Oscar arises from the possibility that if his period-2 partner is an N-type, and has heard that Oscar cheated before, then he will refuse to play with Oscar. To calculate the expected cost, we must calculate the probabilities of the match and of the partner having heard of Oscar’s cheating for each possible location $z$, multiply by the payoff loss $c_2 e^{\theta z}$, and integrate over $z$. The expression for the news probability is different in four different arcs of the circle, as shown in Figure 9.

![Figure 9: Calculation of the expected cost of cheating](image)

22
In the arc between Oscar’s location (O) and that of the period-1 partner he cheated (X), the distance between the general point $Z_1$ and $X$ is $(X - Z_1)$. In the arc extending from $X$ to the point L diametrically opposite O, the distance between the general point $Z_2$ and $X$ is $Z_2 - X$. On the other side of O, upto the point L-X diametrically opposite X, the distance between the general point $Z_3$ and $X$ is $(X + Z_3)$. Finally, between L-X and L, the shorter arc connecting $X$ and the general point $Z_4$ is along the side of the circle away from O, so the distance is $(2L - X - Z_4)$. Then the expression for the expected cost is

$$
\frac{(1 - \epsilon) c_2}{2 \left[ 1 - e^{-\alpha L} \right] / \alpha} \left[ \int_0^X e^{-\alpha Z_1} e^{-\beta(X - Z_1)} e^{\theta Z_1} dZ_1 + \int_X^L e^{-\alpha Z_2} e^{-\beta(Z_2 - X)} e^{\theta Z_2} dZ_2 + \int_0^{L - X} e^{-\alpha Z_3} e^{-\beta(X + Z_3)} e^{\theta Z_3} dZ_3 + \int_{L - X}^L e^{-\alpha Z_4} e^{-\beta(2L - (X + Z_4))} e^{\theta Z_4} dZ_4 \right]
$$

For Oscar’s optimal response to be Honest, this expected cost should exceed the immediate gain. Using the definitions $k = (w_1 - h_1)/c_2$ and $\lambda = \alpha - \theta$, evaluating all the integrals, and simplifying, we get the condition in terms of the function $F(X, L)$ defined in the text.

**Proof of Proposition 1:**

Differentiating the expression in (2),

$$
\frac{\partial F}{\partial X} = \frac{\beta \lambda}{\beta^2 - \lambda^2} \left[ - (e^{-\lambda X} - e^{-\beta X}) + e^{-(\beta + \lambda)L (e^{\beta X} - e^{\lambda X})} \right] - k \theta e^{\theta X} \frac{1 - e^{-\alpha L}}{\alpha}
$$

$$
= - \beta \lambda \frac{e^{\beta X} - e^{\lambda X}}{\beta^2 - \lambda^2} \left[ e^{-(\beta + \lambda)X} - e^{-(\beta + \lambda)L} \right] - k \theta e^{\theta X} \frac{1 - e^{-\alpha L}}{\alpha}
$$

$$
< 0 \quad \text{for } 0 < X < L.
$$

**Proof of Proposition 2:**

When $X = L$, (2) simplifies to

$$
F(L, L) = \frac{e^{-\lambda L} - e^{-\beta L}}{\beta - \lambda} - \frac{k e^{-\lambda L} e^{\alpha L} - 1}{\alpha}
$$

$$
= e^{-\lambda L} \left[ \frac{1 - e^{-(\beta - \lambda)L}}{\beta - \lambda} - k \frac{e^{\alpha L} - 1}{\alpha} \right]
$$

Define $\Phi(L)$ to be the expression in square brackets on the right hand side. The sign of $F(L, L)$ is the same as the sign of $\Phi(L)$. 

23
We have $\Phi(0) = 0$, and 

\[
\Phi'(L) = e^{-(\beta-\lambda)L} - k e^{\alpha L} = e^{\alpha L} \left[ e^{-(\beta+\theta)L} - k \right]
\]

so $\Phi'(0) = 1 - k > 0$ by (P3). Therefore $\Phi(L) > 0$ for $L$ sufficiently small.

If $\beta - \lambda > 0$, then as $L \to \infty$, the first term in the expression for $\Phi(L)$ goes to $1/(\beta - \lambda)$ and the second goes to $\infty$, so $\Phi(L)$ eventually becomes negative. If $\beta - \lambda > 0$, then some rearrangement of terms shows that 

\[
\Phi(L) = e^{\alpha L} \left[ -\frac{e^{-\alpha L} - e^{-(\beta+\theta)L}}{\lambda - \beta} - k \frac{1 - e^{-\alpha L}}{\alpha} \right].
\]

As $L$ becomes large, the first term in the square brackets goes to zero, and the second goes to $-k/\alpha$, so again $\Phi(L)$ becomes negative eventually.

Finally, note that $\Phi'(L) > 0$ if and only if $e^{-(\beta+\theta)L} < k$, that is, $L < -\ln(k)/(\beta + \theta)$. (Since $k < 1$, $\ln(k) < 0$ so the right hand side is positive.) So $\Phi(L)$ changes direction only once. Having become positive for small positive $L$, and becoming negative as $L$ goes to infinity, it can cross zero once and only once. This point defines the unique $L^*$.

**Proof of Proposition 4:**

Letting $L$ go to $\infty$ in (2),

\[
F(X, \infty) = \frac{\beta e^{-\lambda X} - \lambda e^{-\beta X}}{\beta^2 - \lambda^2} - k e^{\theta X} \frac{1}{\alpha} = 0.
\]

Then

\[
F(0, \infty) = 1/(\beta + \lambda) - k/\alpha.
\]

This is positive if and only if $k(\beta + \lambda) < \alpha$; then the $X^*$ defined by $F(X^*, \infty) = 0$ is positive. Otherwise, by continuity of $F$, $F(0, L)$ becomes and stays negative beyond a finite $L$, so $X(L) = 0$ there.

**Proof of Proposition 6:**

From Proposition 2, $L^*$ is defined as the unique positive $L$ satisfying $\Phi(L) = 0$. We also saw in the proof of the proposition that at this point $\Phi(L)$ goes from positive to negative.
values, so $\Phi'(L^*) < 0$. Now take the definition of $\Phi(L)$ in the proof, and notice that it can be written as

$$\Phi(L) = \int_0^L e^{(\alpha - \beta - \theta)z} \, dz - k \int_0^L e^{\alpha z} \, dz = 0.$$ 

To do comparative statics with respect to any parameter, say $k$, differentiate totally:

$$\Phi'(L^*) \frac{\partial L^*}{\partial k} + \frac{\partial \Phi(L)}{\partial k} \bigg|_{L=L^*} = 0.$$ 

Since $\Phi'(L^*) < 0$, the sign of $\partial L^*/\partial k$ is the same as that of the partial derivative of $\Phi(L)$ with respect to $k$. And

$$\frac{\partial \Phi(L)}{\partial k} = -\int_0^L e^{\alpha z} \, dz < 0;$$

therefore $\partial L^*/\partial k < 0$. Similarly,

$$\frac{\partial \Phi(L)}{\partial \beta} = \frac{\partial \Phi(L)}{\partial \theta} = -\int_0^L z e^{(\alpha - \beta - \theta)z} \, dz < 0;$$

therefore $\partial L^*/\partial \beta = \partial L^*/\partial \theta < 0$.

Finally,

$$\frac{\partial \Phi(L)}{\partial \alpha} = \int_0^L z e^{(\alpha - \beta - \theta)z} \, dz - k \int_0^L z e^{\alpha z} \, dz.$$ 

Each of the terms in the expression for $\Phi(L)$ is positive, and the two are equal at $L = L^*$, so dividing each of the terms in the expression for $\partial \Phi(L)/\partial \alpha$ by the corresponding term in the expression for $\Phi(L)$, we see that the sign of $\partial \Phi(L)/\partial \alpha$ at $L^*$ is the same as the sign of

$$\frac{\int_0^L z e^{(\alpha - \beta - \theta)z} \, dz}{\int_0^L e^{(\alpha - \beta - \theta)z} \, dz} - k \frac{\int_0^L z e^{\alpha z} \, dz}{\int_0^L e^{\alpha z} \, dz}.$$ 

Both ratios are weighted averages of $z$ over $[0, L]$. The first has weights proportional to $e^{(\alpha - \beta - \theta)z}$, while the second has weights proportional to $e^{\alpha z}$. So the first gives relatively higher weights to smaller $z$, and is therefore smaller. Thus $\partial \Phi(L)/\partial \alpha$ is negative, and therefore $\partial L^*/\partial \alpha < 0$.

**Proof of Proposition 7:**

Using the workings in the proof of the Lemma, we can write

$$F(X, \infty) = \frac{1}{2} \left[ \int_0^X e^{-\alpha Z_1} e^{-\beta(X-Z_1)} e^{\theta Z_1} \, dZ_1 + \int_X^\infty e^{-\alpha Z_2} e^{-\beta(Z_2-X)} e^{\theta Z_2} \, dZ_2 
+ \int_0^\infty e^{-\alpha Z_3} e^{-\beta(X+Z_3)} e^{\theta Z_3} \, dZ_3 \right] - k \int_0^\infty e^{-\alpha Z} \, dZ.$$
Now $X^*$ is defined by $F(X^*, \infty) = 0$. Differentiating totally with respect to $k$, we have

$$F_X(X^*, \infty) \frac{\partial X^*}{\partial k} + \frac{\partial F(X, \infty)}{\partial k} \bigg|_{X=X^*} = 0.$$  

We know from the proof of Proposition 1 that $F_X < 0$. Therefore the sign of $\partial X^*/\partial k$ is the same as the sign of the second term on the left hand side, which is obviously negative.

An increase in $\beta$ lowers the integrands everywhere in the expression for $F(X^*, \infty)$, so $\partial F(X^*, \infty)/\partial \beta < 0$ and therefore $\partial X^*/\partial \beta < 0$. 


References


———. “On the Interrelations and Economic Implications of Economic, Social, Political and Normative Factors: Reflections from two Late Medieval Societies.” In The Frontiers


