

Social Norms, Local Interaction, and Neighborhood Planning

Matthew Haag* and Roger Lagunoff†

July 11, 2000‡

Abstract

This paper examines optimal social linkage when each individual's repeated interaction with each of his neighbors creates spillovers. Each individual's discount factor is randomly determined. A planner chooses a local interaction network or *neighborhood design* before the discount factors are realized. Each individual then plays a repeated Prisoner's Dilemma game with his neighbors. A *local trigger strategy equilibrium (LTSE)* describes an equilibrium in which each individual conditions his cooperation on the cooperation of at least one "acceptable" group of neighbors. Our main results demonstrate a basic tradeoff in the design problem between suboptimal punishment and social conflict. Potentially suboptimal punishment arises in designs with local interactions since in this case monitoring is imperfect. Due to heterogeneity of discount factors, however, greater social conflict may arise in more connected networks. When residents' discount factors are known to the planner, the optimal design exhibits a cooperative "core" and an uncooperative "fringe." Uncooperative (impatient) types are connected to cooperative ones who tolerate their free riding so that social conflict is kept to a minimum. By contrast, when residents' discount factors are iid, the optimal design partitions individuals into maximally connected *cliques* (e.g., cul-de-sacs). Optimal clique size increases the more patient an individual is likely to be. Finally, if types are correlated, then incomplete graphs with small overlap (e.g., grids) are possible.

JEL Fields: C7, D62, D7.

Keywords: repeated games, local interaction, social norms, neighborhood design, local trigger strategy, spatial design.

*Department of Economics, Georgetown University, Washington DC 20057 USA. Phone: (202) 687-6704. e-mail: haagm@gusun.georgetown.edu.

†Department of Economics, Georgetown University, Washington DC 20057 USA. Phone: (202) 687-1510. e-mail: lagunoffr@georgetown.edu. Web site: www.georgetown.edu/lagunoff/lagunoff.htm

‡We thank seminar participants at American University, Columbia, Cornell, Erasmus, Georgetown, George Washington, Illinois, Iowa, Penn, IUI Stockholm, the NSF/NBER Decentralization Conference, the Midwest Theory and Southeast Theory Conferences, the SED, WEA, and Public Choice Meetings, and the 1999 Summer Econometric Society Meetings for useful comments. We are grateful to Drew Fudenberg and two anonymous referees for very helpful and detailed comments and suggestions.

1 Introduction

It is well known that repeated play can mitigate free rider problems when social spillovers exist. Standard results establish that cooperative outcomes are attainable when individuals are sufficiently patient. These results usually pertain to environments with global interactions — all individuals interact with one another in each repetition of the stage game.¹

Less is known when interactions are local.² Yet, in many instances local rather than global interactions prevail. For example, a neighbor who leaves his porch light on illuminates houses in direct view, but not those located further down the street; an office worker may disrupt the work of those in adjacent cubicles, but not in non-adjacent ones; a gas station that lowers its price attracts customers away from the station across the street but, perhaps, not from the one two blocks away.

These local interaction settings share the characteristic that each person interacts with only a subset of the relevant population. Moreover, the interaction need not be transitive; one’s strategic “neighbors” may not be the same as one’s “neighbor’s neighbors.” This paper examines repeated play in such settings. Specifically, our goal is to compare consequences of repeated play across different spatial designs. It turns out that some designs are more conducive than others to socially desirable outcomes. We therefore address the question of the *optimal spatial or neighborhood design* when free rider problems are localized.

To focus on spillovers and free rider problems, this paper describes a design problem in which each individual in a large society plays a repeated Prisoners Dilemma game with his neighbors. An individual’s stage payoff is the sum of the payoffs from each neighbor interaction. A social planner must choose a *neighborhood design*, a local interaction system represented by an undirected graph. Each individual only interacts with, and observes behavior of, those with whom he is linked.

The planner makes his choice knowing only the joint distribution of “types” of individuals. A useful analogy is to that of a city planner who knows something about the aggregate population characteristics, but does not know the identities of the residents who move in after the houses are built. We assume an individual’s type is his discount factor. Randomly determined discount factors introduce heterogeneity in the population.

Rather than study the entire, complicated set of sequential equilibrium interactions, we limit attention to a particular subclass. We examine norms of conduct that rely on simple punishments to enforce social cooperation. The concept of a *local trigger strategy equilibrium (LTSE)* is introduced to describe a “trigger strategy” equilibrium in which each individual chooses to condition his cooperation on the continued cooperation of at least one

¹See, for example, Fudenberg and Maskin (1986). Some results also pertain to environments with random, pairwise matching. See Kandori (1991), Ellison (1994), and Okuno-Fujiwara and Postlewaite (1995).

²Related models are discussed in Section 5.

“acceptable” group of neighbors. For instance, an individual living on a traditional street grid may choose to cooperate as long as either of his two adjacent neighbors continues to do so. However, if both neighbors ever choose to play uncooperatively then the individual permanently reverts to uncooperative behavior himself. As in other repeated PD games, an individual’s cooperation requires a certain threshold level of patience.

Certainly, some generality is lost in restricting attention to these trigger strategies. Nevertheless, we show that LTSE are simple, easily followed norms which exhibit a large variety of “in-equilibrium” free riding, selective enforcement, and other social phenomena that emerge naturally from the heterogeneous agent, local interaction environment. We focus on optimal LTSE — those that maximize the planner’s criterion in any neighborhood design.

The main results of this paper demonstrate a basic tradeoff in the design problem between suboptimal punishment and social conflict. Potentially suboptimal punishment arises in designs with local interactions since these interactions have private monitoring. Due to heterogeneity of discount factors, however, greater social conflict may arise in more connected networks. This tradeoff is illustrated in some simple examples in Section 2. The benchmark model is described in Section 3.

Section 4 studies the case in which all groups of connected individuals are maximally connected. Maximally connected graphs, or *cliques*, conform to the case where (1) the actions of each neighbor are consequential to all others, and (2) these actions are observable to the others. For example, cul-de-sacs (circular streets in which each house is in plain view of the others) have this property. Cliques are of particular interest since maximal linkage is assumed implicitly in standard repeated games with perfect monitoring. We characterize the planner’s problem when her choices are restricted to cliques. Optimal clique size varies depending on the payoff and distributional characteristics. We show that when types are distributed iid, optimal size is increasing in first order stochastic increases in the type distribution. Hence, the higher the likelihood of patient types, the larger is the optimal size of a clique. What makes the result nontrivial is that independence of types does not imply independence of one’s likelihood of cooperation. Higher threshold discount factors are necessary for cooperation the smaller the number of other cooperators in one’s neighborhood.

The general design problem is taken up in Section 5. The first main result characterizes the “full information” solution, i.e., the solution when the distribution over discount factors is degenerate. A special case of this is the standard assumption of homogeneous discount factors.³ Not surprisingly, when everyone is sufficiently patient, the largest possible clique is uniquely optimal. The problem is more interesting when discounting is bounded away from 1. Here we show that cliques need not be optimal. Generally, the solution is shown to exhibit a cooperative “core” and an uncooperative “fringe.” Sufficiently patient neighbors

³With some notable exceptions (see Harrington (1989) and Lehrer and Pauzner (1999)), the individuals in repeated game models are assumed homogeneous with respect to rates of time preference.

(“cooperative types”) are maximally connected to one another. However, impatient or “uncooperative” types are also connected to certain individuals in the cooperative group who are able to tolerate some free riding. Those who are both cooperative yet comparatively intolerant of free riders are connected to fewer uncooperative neighbors. In this sense, the optimal design minimizes the degree of social conflict.

When the type distribution is not degenerate, the planner may not be able to prevent social friction *ex post*. Unlike the full information solution, when the type distribution is iid the optimal design is shown to partition individuals into identical cliques. The intuition for this result is that since types are iid, optimal neighborhood designs will generally be symmetric. That is, the number of links to any individual should be the same. But symmetric graphs with low connectivity (small numbers of links per person relative to the maximum of $n - 1$ in a component of size n) provide more incentives to free ride. The reason is that sanctions against free riding may not be credible if too many of the sanctioner’s neighbors do not observe the defection. This is illustrated in Section 2. Low connectivity results in imperfect monitoring. Maximal connectivity provides perfect monitoring, and therefore provides maximal incentives to cooperate.

When some correlation in the joint distribution over types exists, optimal designs with grid-like features are possible. Without independence, one person’s likelihood of realizing a discount factor above a particular threshold may negatively or positively correlate with another’s. We construct examples where grid designs are optimal.

Section 6 concludes with a review of related literature and a discussion of extensions and modifications. Finally, Section 7 is an appendix with proofs of some of the results.

2 An Example

The two figures below correspond to distinct “neighborhood designs.” In each, three individuals are represented abstractly as nodes in a graph. The links express both the presence of potential spillovers and the information flows. In Figure 1a, the individuals are maximally connected in a *clique*, while in 1b the link between Neighbor 2 and Neighbor 3 is absent. In the latter, neither Neighbor 2 nor Neighbor 3 observes the behavior of the other.⁴ One interpretation is that in Figure 1a, these three live in houses arranged in a *cul-de-sac*, while in Figure 1b, they live on a street grid.

In each link, the interaction each period is described by the PD game in Figure 2 below. Each person can choose one of two actions, “*C*” or “*D*”. Action “*C*” corresponds to “cooperative” behavior. In the case where the local interaction conforms to residential settings,

⁴If, for instance, a neighbor enjoys viewing another neighbor’s flower bed, then the spillover coincides with his observations of his neighbor.

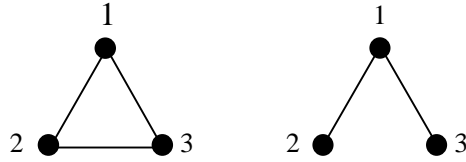


Figure 1:

the cooperative action may be mowing the lawn, taking in one’s garbage can at the end of the day, leaving an outdoor light on, or volunteering to maintain a communal garden. Conversely, action “ D ” is defined as the “deviant” or “uncooperative” action. In each of these examples, an individual’s cooperative action confers external rewards to others.

	C	D
C	c, c	$-\ell, d$
D	$d, -\ell$	$0, 0$

Figure 2

We assume $d > c > 0 > -\ell$. Payoff c is the gain from mutual cooperation, d the gain from “deviant” behavior, and ℓ the loss from being cheated. We also assume that $2c > d - \ell > 0$ so that mutual cooperation is Pareto undominated by payoffs in the convex hull of the payoff set,⁵ and the aggregate payoff is higher if a cooperative individual “tolerates” uncooperative behavior by another.

While Figure 2 describes the payoffs from a single interaction between two individuals, we examine activities with spillovers such as those mentioned above, whereby an individual cannot isolate his behavior toward one neighbor from his behavior toward other neighbors. In each period, an individual’s choice of “ C ” or “ D ” applies to all his links simultaneously.

In standard analysis of repeated PD, lower bounds on the discount factor δ_i for each individual $i = 1, 2, 3$ determine when full cooperation, i.e., all neighbors choose “ C ”, is an equilibrium of the repeated game. To find this bound, one may examine a perfect equilibrium in which the “grim trigger strategy” — defection is met with permanent reversion to the Nash equilibrium — is used. If other neighbors adopt this strategy, then each individual i will choose “ C ” if $c > (1 - \delta_i)d$, i.e., the payoff from permanent cooperation outweighs the one shot gain from defection. Full cooperation is therefore an equilibrium if

$$\delta_i \geq \frac{d - c}{d}, \quad i = 1, 2, 3. \quad (1)$$

⁵Any payoff in convex hull can be arbitrarily approximated by a time averaged pure behavior strategy.

In both designs, Inequality (1) is necessary for full cooperation to constitute a perfect equilibrium. In the clique, it is also sufficient. Since the interaction there is global, it is no different from a standard three-person repeated game. By contrast, in the grid, there is potentially an additional “perfection” constraint for Neighbor 1. When the grim trigger punishments are used, this “perfection” constraint in the grid requires that if, for instance, Neighbor 3 defects to “ D ”, then Neighbor 1 must retaliate by choosing “ D ” himself, rather than tolerate 3’s behavior, even for just one period. However, by punishing Neighbor 3, he threatens his relationship with Neighbor 2 who did not observe the deviation. This constraint is satisfied if $(1 - \delta_1)d \geq (1 - \delta_1)(c - \ell) + \delta_1(1 - \delta_1)d$. Rewriting this inequality yields an upper bound on δ_1 ,

$$\delta_1 \leq \frac{d - c}{d} + \frac{\ell}{d} \tag{2}$$

If $1 > \delta_1 > \frac{d - c}{d} + \frac{\ell}{d}$ then an excessively patient Neighbor 1 using this strategy will not punish a single deviator on either side.⁶ Instead, he will tolerate uncooperative behavior by one of his two neighbors. In such a case, potential gain to “containment” of the bad behavior on one side outweighs the gain to retaliation when a neighbor chooses “ D ”.

While the constraint in (2) describes the incentive to credibly *punish* an out-of-equilibrium deviation, the violation of (2) describes the in-equilibrium incentive to *tolerate* one uncooperative neighbor. Inequalities (1) and (2) therefore define three relevant intervals of discount factors. Low types satisfy $\delta_L < \frac{d - c}{d}$, and hence, will never cooperate. “Tolerant” types satisfy $\delta_T \geq \frac{d - c}{d} + \frac{\ell}{d}$, meaning that they are patient enough, not only to cooperate themselves, but will also tolerate uncooperative behavior from one other neighbor. Finally, “intolerant” types are cooperative but intolerant of uncooperative behavior, i.e., $\frac{d - c}{d} \leq \delta_I < \frac{d - c}{d} + \frac{\ell}{d}$.

Suppose, first, that all individuals are exceedingly patient, i.e., $\delta_i = \delta_T$, $i = 1, 2, 3$. It is easy to see that the full-cooperation equilibrium exists in the clique. However, because (2) is violated for Neighbor 1, this full-cooperation equilibrium does not exist for the grid. The reason is that Neighbor 3 can choose “ D ” with impunity (alternatively, if Neighbor 3 cooperates then Neighbor 2 can choose “ D ” with impunity). Since Neighbor 1 will tolerate such a defection, Neighbor 3 has no reason to continue to cooperate. In this sense, the clique has higher aggregate welfare than the grid.

Now suppose $\delta_1 = \delta_T$, $\delta_2 = \delta_I$, $\delta_3 = \delta_L$. Since Neighbor 3 is a low type, full cooperation is not possible in any spatial arrangement. However, the following is a simple, stationary sequential equilibrium in the grid. Neighbor 3 chooses “ D ”, Neighbor 1 chooses “ C ”, thereby tolerating Neighbor 3’s bad behavior, and Neighbor 2 choose “ C ”.

⁶If discount factors are homogeneous, then Ellison (1994) describes an ingenious way of constructing sequential equilibria for δ close to one that simulates the situation of smaller discount factors. Roughly, given an equilibrium for a small discount factor, his construction replicates this equilibrium on every N th period of the repeated game for an appropriate choice of N . Consequently, equilibria exist for which the constraint (2) does not apply. However, since this method uses the same re-normalization for everyone’s discount factor, it may not work when discounting is sufficiently heterogeneous. We do not know if there exists a similar result for a heterogeneous world.

By contrast, this “partial cooperation” equilibrium does not exist in the clique. While the link between the cooperating neighbors, Neighbors 1 and 2, poses no problem, the link between Neighbors 2 and 3 is problematic. Neighbor 3’s behavior, while tolerable to Neighbor 1, is intolerable to Neighbor 2. Discount factor δ_2 satisfies both (1) and (2) which means that Neighbor 2 is patient enough to cooperate in equilibrium, but not patient enough to tolerate a neighbor choosing “ D ”. Therefore, in a clique Neighbor 2 will only respond to 3’s choice by choosing “ D ”, leaving Neighbor 1 no alternative but to choose “ D ”. Since all three choose “ D ” in the clique, the grid has higher aggregate welfare. Note also that the assumption $d - \ell > 0$ implies that the grid is better than a design that disconnects Neighbor 3 altogether.

The examples highlight an important tradeoff. If all individuals are patient, then structures with a high degree of social interaction, such as large cliques, are preferable. However, when uncooperative types are present, then designs that limit social interaction among certain neighbors may be preferable. In the grid example, Neighbor 1, the more tolerant of the two “cooperative” types, effectively “shields” behavior of the uncooperative Neighbor 3 from the less tolerant Neighbor 2. This tradeoff is relevant to a planner even if she does not have precise information about individuals’ discount factors.

3 The Model

3.1 Neighborhood Designs

A set M of individuals with $m = |M|$ play an infinitely repeated game with network externalities. At the beginning of time $t = 0$, a residential planner chooses a local interaction system that determines who interacts with whom. We refer to this structure as a *neighborhood design* which is described as a collection of subsets $N = (N_1, \dots, N_m)$. For each individual i , the set N_i is the collection of individuals with whom i interacts each period. We refer to this subset as “ i ’s neighborhood”. Hence, $j \in N_i$ means that j is i ’s neighbor. We assume that $i \in N_i$ (N_i is self inclusive). We also assume that the relation is symmetric so that $j \in N_i$ iff $i \in N_j$.

The neighborhood design describes a repeated game of private monitoring. Specifically, $i \in N_j$ means not only that i and j directly interact, but also that i and j each observe the history of the other’s behavior. Conversely, $i \notin N_j$ implies that i does not observe j ’s behavior.⁷

As before, we assume that the game in Figure 2 describes an individual’s interaction with each of his neighbors. Let a_i^t denote the action taken by individual i in period t .

⁷See Section 6 for extensions that break this link.

Let $a_{N_i}^t = (a_j^t)_{j \in N_i}$ denote the behavior of everyone in i 's neighborhood (including, by definition, i himself). Individual i 's t -period *personal history* is defined as the list $h_i^t = (a_{N_i}^1, a_{N_i}^2, \dots, a_{N_i}^{t-1})$. A standard notation denotes H_i as the set of all t -period personal histories for $i \in M$.

Individual i 's dynamic payoff is the discounted sum $\sum_t (1 - \delta_i) \delta_i^t u_i(a_{N_i}^t)$ where δ_i is i 's discount factor, and u_i is i 's stage payoff function. Individual i 's payoff at date t is the sum of the payoffs from each interaction. Hence, if q other individuals in neighborhood N_i play “C” at time t then i 's temporal payoff is $qc - (n_i - 1 - q)\ell$ if he plays “C”, and is qd if he plays “D”. This specification captures both negative scale effects of congestion if most neighbors choose “D”, and positive scale effects of spillovers if most neighbors choose “C”.

We assume that at date $t = 0$, before play begins, each individual's discount factor is randomly determined. Let G denote the joint distribution over vectors $\delta = (\delta_1, \dots, \delta_m)$. Once a vector δ is realized, it is common knowledge to everyone at the start of play. Each individual then chooses a behavior strategy f_i in the ensuing local interaction game. For each personal history h_i , we write $f_i(h; \delta) \in \{C, D\}$ to denote i 's action as a function of history h given a parameter vector of realized discount factors δ .⁸ Societal behavior in the interaction game can now be completely described by a profile $f = (f_i)_{i \in M}$ of strategies. Clearly, the “uncooperative” profile in which everyone plays “D” regardless of history is a sequential equilibrium.

The model differs from standard repeated games with perfect monitoring in two respects. First, local interaction implies that a neighbor's behavior may not be observed by one's other neighbor. Second, the dispersion of discount factors δ_i introduces heterogeneity across individuals. The presence of any number of impatient individuals implies that the Folk Theorem will not generally apply. In such cases, certain neighborhoods may tolerate some degree of in-equilibrium “cheating.”

A planner who knows G must choose a neighborhood design before the realization δ . If, following the planner's chosen neighborhood design $N = (N_1, \dots, N_m)$, a strategy profile f is anticipated, then the planner's criterion is the expected discounted sum

$$E \left[\frac{1}{m} \sum_{i \in M} \sum_t (1 - \delta_i) \delta_i^t u_i(\tilde{a}_{N_i}^t(f)) \right], \quad (3)$$

where E is the expectation operator taken with respect to distribution G , and $\tilde{a}_{N_i}^t(f)$ is the action profile of i 's neighbors induced by anticipated profile f . The criterion (3) presumes equal treatment of all individuals ex ante.⁹ Criterion (3) attains a maximum value of

⁸Only pure strategies are considered in this analysis.

⁹An editor rightly pointed out to us that by using the welfare criterion (3) we implicitly rule out the possibility of interpreting heterogeneous discounting as arising from differing frequencies of play across individuals. Instead, we interpret the differing discount factors as coming from differences in the pure rate of time preference; some individuals are simply assumed to be more impatient than others.

$(m-1)c$ since there are $m(m-1)/2$ possible linked pairs, and $2c$ is the maximal total payoff for each link. The maximal value occurs when everyone permanently cooperates and every pair of individuals is linked.

Unfortunately, two issues arise that require render criterion(3) inadequate for our purposes. First, if the planner does not directly influence the choice of f , then any design can maximize (3) provided that the residents play the uncooperative equilibrium following any other chosen design.¹⁰ To rule out designs which are supported by this “punishment-by-beliefs” scenario, we ultimately restrict attention to sequential equilibrium profiles f that maximize (3) given N . We call such a profile f^* *optimal in design N* . Let $W(N)$ denote the value of (3) in the optimal equilibrium f^* . A design N is *optimal* if it solves:

$$\max_N W(N) \tag{4}$$

Second, for tractability we consider solutions to (4) that ignore integer and remainder problems. For example, suppose that the optimal design called for a replication of finite component graphs, each with r individuals. To implement this design would require a population size divisible by r . For our purposes we simply assume such divisibility. While this entails some loss of generality, our solutions will approximate the actual solutions when the population relative to the solution size is large enough or when the solution leaves a small enough remainder.

3.2 Local Trigger Strategy Equilibria

Here we introduce a type of stationary, “trigger strategy” equilibrium for repeated play in graphs. To facilitate our definition, some definitions and notation are needed.

First, for each individual $i \in M$ and for each personal history h_i^t let $N_i(h_i^t)$ denote the set of neighbors that have never chosen the deviant action, “ D ”, in the past, i.e.,

$$N_i(h_i^t) = \{j \in N_i : a_j^\tau \neq D, \forall \tau < t\}.$$

Notice that $N_i(h_i^1) = N_i$, i.e., everyone in the neighborhood is presumed cooperative in the null history. Next, a collection, $\mathcal{T} \subset 2^M$, of sets is *comprehensive* if for any nonempty set $S \in \mathcal{T}$, if $S' \supseteq S$, then $S' \in \mathcal{T}$. Comprehensive collections are those that are closed under the taking of supersets of nonempty sets.

¹⁰Specifically, suppose there are two designs, N and N' . In design N all residents choose D (the uncooperative equilibrium), even though other equilibria exist. In design N' , residents choose an equilibrium that admits some social cooperation even though the uncooperative equilibrium always exists. While design N' then yields a higher value according to (3), the comparison is biased by the selective choice of equilibrium in each.

Definition 1 A local trigger strategy for individual i in the neighborhood design N is a strategy f_i satisfying: for each vector δ , there exists a comprehensive collection $\mathcal{T}_i^\delta \subseteq 2^{N_i}$ of subsets of neighbors of i such that for each personal history h_i ,

$$f_i(h_i; \delta) = \begin{cases} C & \text{if } N_i(h_i) \in \mathcal{T}_i^\delta \setminus \{\emptyset\} \\ D & \text{otherwise.} \end{cases}$$

In words, a local trigger strategy in a neighborhood design is one in which an individual agrees to cooperate if and only if the permanent cooperators in i 's neighborhood is “acceptable” according to \mathcal{T}_i . Local trigger strategies require each person to bind his behavior to a judicious, though not necessarily unique, selection of trustworthy members of his community. Comprehensivity guarantees that reciprocity is (weakly) increasingly likely, the larger is the set of cooperators. It also guarantees that “ D ” is an absorbing action for each individual. Since $N_i(h_i)$ can never increase over time, once $N_i(h_i) \notin \mathcal{T}_i^\delta \setminus \{\emptyset\}$, no subsequent cooperating set can be included in $\mathcal{T}_i^\delta \setminus \{\emptyset\}$.

Hereafter, we associate resident i 's local trigger strategy with the set \mathcal{T}_i^δ and refer to it as i 's collection of “trigger sets.” The collection $\mathcal{T}_i^\delta = \{\emptyset\}$, for example, corresponds to the strategy, “always play D .” The trigger set $\mathcal{T}_i^\delta = \{S' : S' \supseteq S\}$ for some $S \subset N_i$ corresponds to the strategy, “play C as long as everyone in S continues to play C ; play D otherwise.” In this case, individual i 's cooperation depends on the continued cooperation of everyone in set S .¹¹ Such a strategy may be contrasted with the trigger set $\mathcal{T}_i^\delta = \{\{j\}, \{k\}, \{j, k\}\}$ which corresponds to: “play C if either j or k continue to play C ; play D otherwise.”

In Figure 3 below, we describe an example of a sequential equilibrium in which each individual uses a local trigger strategy. Using the three types, δ_L , δ_T and δ_I defined in the previous Section, suppose that Neighbor 3 in Figure 3 is tolerant of a single free rider, i.e., $\delta_3 = \delta_T$, while all others are cooperative but intolerant, i.e., $\delta_i = \delta_I$ for all $i \neq 3$. Now define: $\mathcal{T}_i^\delta = \{N_i\}$ for each $i \neq 3$, while $\mathcal{T}_3^\delta = \{\{2, 3\}, \{3, 4\}, \{2, 3, 4\}\}$. While each of the others requires cooperation from everyone in his neighborhood, Neighbor 3 only requires it from himself and either of his two neighbors. Despite Neighbor 3's failure to punish each neighbor individually, both Neighbors 2 and 4 will cooperate since sufficient deterrence is provided by their end-point neighbors.

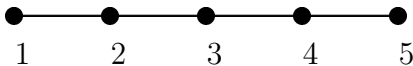


Figure 3:

¹¹While the supersets of S are also acceptable, only the cooperation of members of S is essential for i 's continued cooperation.

Certainly, local trigger strategies entail some loss of generality in the design problem. Since nonstationary strategies are not considered, a relatively impatient neighbor plays “ D ” permanently. The community might be better off if an impatient type could choose “ C ” every n th period. However, to be successful in an environment with heterogeneous discounting, time varying strategies such as this may require long recall and carefully calibrated sanctioning tailored to the realized type of each individual. By contrast, local trigger strategies are quite simple. They admit a very simple partition of private histories into two states, “cooperative” and “punishment/free riding,” the latter being absorbing. Despite their simplicity, local trigger strategies will be shown to admit a large degree of “in-equilibrium” free riding, selective enforcement, and other interesting social interactions when spillovers are present.

Definition 2 *A local trigger strategy equilibrium (LTSE) in neighborhood design N is a collection of trigger strategy profiles, $\mathcal{T} = (\mathcal{T}_i^\delta)_{i \in N, \delta \in [0,1]^m}$, that comprises, for each δ , a sequential equilibrium. An optimal LTSE is an LTSE that maximizes the social welfare criterion (3) in the class of LTSE.*

As Section 2 showed, for some realizations of δ , LTSE may entail that only a subset of individuals cooperate in equilibrium. Clearly, the set of LTSE is always nonempty as it contains the “uncooperative equilibrium” $\mathcal{T}_i^\delta = \{\emptyset\}$ for each i and each δ . (Note that by our assumption that $N_i(h_i^1) = N_i$, i.e., that everyone is presumed cooperative after the null history, an individual who chooses “ D ” in equilibrium will choose $\mathcal{T}_i^\delta = \{\emptyset\}$. Hence, $\mathcal{T}_i^\delta \neq \{\emptyset\}$ implies that i cooperates in equilibrium.) In all that follows, we assume that the planning problem is solved assuming that optimal LTSE are used (where optimality is restricted in the sense that it is relative to the set of LTSE). The notion of *optimal design* is relative to this class.

4 Optimal Cliques

We first consider a special case in which the Planner is restricted to neighborhood designs consisting of partitions into maximally connected neighborhoods or “cliques.” These are useful objects to study since they correspond to interactions in the standard repeated game model. They also occur naturally in many real environments. For example, a neighbor who volunteers for a “Neighborhood Watch Association” patrols the entire community.

Assume then that the Planner chooses N from among the set of cliques of size no greater than m to solve (4). In this case, the question of structure reduces to one of size. To understand the determinants of optimal clique size, one must first characterize incentive constraints implied by optimal LTSE, then use this characterization to construct the planner’s criterion as a function of size. We obtain an explicit construction only for the

case where types (discount factors) are distributed iid. In the subsequent Section we prove that cliques are optimal among all designs precisely when types are iid.

The iid assumption on types allows us to focus on a single clique and drop “ i ” subscripts for convenience. Formally, we may write the planner’s criterion as $W(n)$ where n is the size (number of individuals) in the clique. The optimal clique size is an integer n that maximizes

$$W(n) = \sum_{s=0}^n p(s|n)v(s; n) \quad (5)$$

where $p(s|n)$ denotes the probability that s individuals are cooperative in an optimal LTSE in a clique of size n , while $v(s; n)$ is the aggregate value when s individuals are cooperative in an n -clique, and is given by

$$v(s; n) = \frac{1}{n} \left[s \left((s-1)c + (n-s)\frac{d-\ell}{2} \right) + (n-s)(s-1)\frac{d-\ell}{2} \right]$$

To maximize (5), a characterization of p is required. First, observe that since all neighbors are connected to one another in any clique, deviations are common knowledge in the group. If one neighbor deviates from prescribed cooperation, he can be sanctioned by all neighbors, each knowing that all other neighbors will enforce sanctions. There is perfect coordination of sanctions among all neighbors. Therefore, the equilibrium level of cooperation in any LTSE is determined by the set of neighbors that satisfy a collection of simple, single threshold, incentive constraints. Expecting that a deviation is met with immediate “grim trigger” sanction by all neighbors, an individual will cooperate in equilibrium if $(s-1)c - (n-s)\ell \geq (1-\delta_i)(s-1)d$ or

$$\delta_i \geq \min \left\{ 1, \frac{d-c}{d} + \frac{\ell(n-s)}{d(s-1)} \right\} \quad (6)$$

Notice that (6) is a generalization of the Inequalities (1) and (2). Clearly, the constraint is less restrictive the larger is the number, $s-1$, of cooperating neighbors.

Using Inequality (6), the type space of discount factors can be segmented into intervals, each consisting of types that tolerate precisely certain numbers of cooperating neighbors. For the clique of size n , let $H_{nq} \subseteq [0, 1]$ denote the set of types (discount factors) for each individual which cannot tolerate q cooperating neighbors, but can tolerate $q+1$ or more cooperative neighbors where $q < n$. When $q = n-1$ then the individual is an impatient type. Using Inequality (6), one can verify that $H_{nn-1} = [0, \frac{d-c}{d})$, and

$$H_{nq} = \left[\min \left\{ 1, \frac{d-c}{d} + \frac{\ell[n-1-(q+1)]}{d(q+1)} \right\}, \min \left\{ 1, \frac{d-c}{d} + \frac{\ell(n-1-q)}{dq} \right\} \right) \\ \text{if } q = 0, \dots, n-2.$$

Recall that G is the joint distribution on δ . Let G_i denote the marginal distribution on δ_i . With iid types, we assume that each identical marginal G_i admits a density g so that $g(H_{nq}) \equiv \int_{H_{nq}} dG_i$ denotes the probability of H_{nq} . Then, let

$$G_{ns} = \sum_{q=s}^n g(H_{nq}) = G_i \left(\min \left\{ 1, \frac{d-c}{d} + \frac{\ell(n-s)}{d s} \right\} \right)$$

In words, G_{ns} is the probability that an individual i is too impatient to cooperate when s of his neighbors are cooperative.

Our construction of p is nearly complete. Given any set of k individuals, we will say that a vector $r = (r_1, \dots, r_k)$ with $r_i \in \{0, 1, \dots, k-1\}$ for each i is *k-admissible* if there exists a permutation $\theta : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$ such that the permuted vector r_θ satisfies $r_\theta \geq (0, 1, 2, \dots, k-1)$. Let $\mathcal{A}(k)$ denote the set of k -admissible vectors. An admissible point r is an integer-valued vector which, under some permutation, is bounded below by a stepwise ascending vector. For example, $r = (0, k-1, k-1, \dots, k-1)$ is admissible. However, $r = (0, 0, k-1, k-1, \dots, k-1)$ is not.

Lemma 1 *The equilibrium distribution p on the number of cooperators in an n -clique is given by*

$$p(s|n) = \binom{n}{s} (1 - G_{n, s-1})^s \sum_{r \in \mathcal{A}(n-s)} \prod_{j=1}^{n-s} g(H_{n, s+r_j})$$

The proof is in the Appendix. Probability p describes a particular type of correlated binomial distribution (“success” likelihoods may be correlated). Like the standard binomial, this distribution is unimodal.¹² From the Lemma an explicit solution to (5) can be computed. Additionally, we can prove that optimal clique sizes (weakly) increase in first order stochastic shifts in G .¹³

Theorem 1 *Suppose that types are iid distributed. Suppose that the marginal distribution G_i^1 first order stochastically dominates marginal distribution G_i^2 . Then the optimal sized clique under G_i^1 is at least as great as that under G_i^2 .*

Proof Consider a clique of size n . Let W^1 and W^2 denote values of the planner’s criterion

¹²For a description and properties of correlated binomial distributions, see Johnson, Kotz, and Kemp (1992), pp. 148-150.

¹³A distribution G^1 first order stochastically dominates G^2 if for each x , $G^2(x) > G^1(x)$.

under distributions G^1 and G^2 , resp. Observe that

$$\begin{aligned} v(s; n) &= \frac{1}{n} \left[s \left((s-1)c + (n-s) \frac{d-\ell}{2} \right) + (n-s)(s-1) \frac{d-\ell}{2} \right] \\ &= \frac{1}{n} \left[s(s-1)c + (n-s)(2s-1) \frac{d-\ell}{2} \right] \end{aligned}$$

One can verify that v is strictly increasing in s . Moreover, v is increasing when s increases while adjusting n so that s/n is held constant. Now for arbitrary n , examine the difference:

$$\begin{aligned} \Delta_s(n) &\equiv p^1(s|n) - p^2(s|n) \\ &= \binom{n}{s} \left[(1 - G_{n \ s-1}^1)^s \sum_{r \in \mathcal{A}(n-s)} \prod_{j=1}^{n-s} g^1(H_{n \ s+r_j}) \right. \\ &\quad \left. - (1 - G_{n \ s-1}^2)^s \sum_{r \in \mathcal{A}(n-s)} \prod_{j=1}^{n-s} g^2(H_{n \ s+r_j}) \right] \end{aligned} \quad (7)$$

Clearly, $\sum_s \Delta_s(n) = 0$. Equation (7) implies that $\Delta_s(n) \geq 0$ iff

$$\left(\frac{1 - G_{n \ s-1}^1}{1 - G_{n \ s-1}^2} \right)^s \geq \frac{\sum_{r \in \mathcal{A}(n-s)} \prod_{j=1}^{n-s} g^2(H_{n \ s+r_j})}{\sum_{r \in \mathcal{A}(n-s)} \prod_{j=1}^{n-s} g^1(H_{n \ s+r_j})}$$

Then $\Delta_s(n)$ is increasing in s , and $\Delta_s(n) < 0$ when $s = 0$, and $\Delta_s(n) > 0$ when $s = n$. This gives us a standard First Order Stochastic Dominance property that $\sum_{s \leq \bar{s}} \Delta_s(n) \leq 0$, for any $\bar{s} \leq n$. This, in turn, implies $W^1(n) \geq W^2(n)$ for any n .

Hence, to prove the result it suffices to show that for the range of n where W^2 is increasing, W^1 must be increasing as well. If this is the case, then $n^1 \geq n^2$ where $n^1 = \arg \max_n W^1(n)$ and $n^2 = \arg \max_n W^2(n)$, respectively.¹⁴

Let $n > n'$. It suffices then to show that $W^2(n) - W^2(n') > 0$ implies $W^1(n) - W^1(n') > 0$. In particular, we show that $W^1(n) - W^1(n') > W^2(n) - W^2(n')$ or

$$W^1(n) - W^2(n) > W^1(n') - W^2(n') \quad (8)$$

Inequality (8) is equivalent to

$$\sum_{s=0}^n [p^1(s|n) - p^2(s|n)]v(s; n) > \sum_{s=0}^{n'} [p^1(s|n') - p^2(s|n')]v(s; n') \quad (9)$$

¹⁴At least generically, we conjecture that a maximizer n of W is unique, though we have been unable to show it. However, if there are multiple solutions, then our proof establishes that for at least one solution n^2 of W^2 and one solution n^1 of W^1 , $n^1 \geq n^2$.

Hence, we must verify (9) to prove the result. For every $s' = 0, \dots, n'$, let s be the smallest integer such that $s/n \geq s'/n'$. That is, s/n is the closest approximation (from above) of the proportion of cooperators in the n -clique to s'/n' in the n' -clique. Observe that for each s' , a unique s exists which serves the purpose. Let ψ denote the corresponding map $s' \mapsto s$. If $s = \psi(s')$, then $s > s'$, $v(s; n) > v(s', n')$, and by (7), $\Delta_s(n) > \Delta_{s'}(n')$. These facts imply

$$\begin{aligned} \sum_{s=0}^n [p^1(s|n) - p^2(s|n)]v(s; n) &\geq \sum_{s'=0}^{n'} [p^1(\psi(s')|n) - p^2(\psi(s')|n)]v(\psi(s'); n) \\ &> \sum_{s'=0}^{n'} [p^1(s'|n') - p^2(s'|n')]v(s'; n') \end{aligned} \tag{10}$$

which establishes (9). \diamond

Unfortunately, the result asserts only a weak inequality. This is due to the fact that $W(n)$ is an integer program, and so arbitrarily small changes in G may have no effect on the solution. To see how optimal clique size may vary, suppose that $(1 - G_i(\frac{d-c}{d})) = 1$. Then all individuals are commonly known to be cooperative and so $W(n) = (n-1)c$. In this case, optimal clique size is unbounded (see also Theorem 3 in the next Section). By contrast, if $G_i(\frac{d-c}{d}) = 1$, then all individuals are commonly known to be uncooperative. Therefore, $W(n) = 0$ for all n .

If there is any positive mass above $\frac{d-c}{d}$ then the planner's payoff cannot be zero. Nevertheless, one can construct distributions that bound optimal clique size. For example, let α satisfy: $0 < \alpha < G_i(\frac{d-c}{d})$ and $G_i(\frac{d-c}{d} + \frac{\ell(1-\alpha)}{d}) = 1$. This last condition implies that no individual will tolerate less than fraction α cooperative neighbors. With this distribution,

$$W(n) < 4c \sum_s s p(s|n) = 4c \sum_{s \geq \alpha n} s p(s|n) < 4c \sum_{s \geq \alpha n} s \binom{n}{s} \left(1 - G_i\left(\frac{d-c}{d}\right)\right)^s G_i\left(\frac{d-c}{d}\right)^{n-s} \tag{11}$$

The last inequality follows from the fact that in the standard binomial distribution with success rate $(1 - G(\frac{d-c}{d}))$, types above $(d-c)/d$ are presumed cooperative, while in the actual distribution p , some types above that threshold may not cooperate. Hence, the standard binomial distribution with success rate $(1 - G(\frac{d-c}{d}))$ first order stochastically dominates the actual distribution p . Since $\alpha > (1 - G(\frac{d-c}{d}))$, the right hand side of the Inequality in (11) converges to zero as $n \rightarrow \infty$.¹⁵ Hence, $\lim_{n \rightarrow \infty} W(n) = 0$, and so optimal clique size is bounded.

¹⁵Given the assumptions on α , the right tail, $\sum_{s \geq \alpha n} \binom{n}{s} (1 - G_i(\frac{d-c}{d}))^s G_i(\frac{d-c}{d})^{n-s}$, is $o(\frac{1}{a^n})$ for some constant $a > 1$.

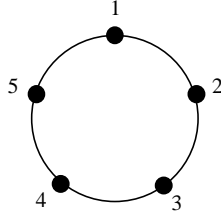


Figure 4:

5 Optimal Designs: The General Case

5.1 A Problem with Local Trigger Strategies

We now turn to the general design problem. As with cliques, a characterization of optimal LTSE is required before proceeding with the planner’s problem. Unfortunately, when designs are incompletely connected, the restriction implied by sequential equilibria in local trigger strategies can be unduly harsh.

In the example in Figure 4 above, each Neighbor has the same discount factors as in Figure 3 from Section 3. The difference is that Neighbors 1 and 5 are linked, closing the graph. With this change, the only LTSE is one in which all individuals play “ D ”! The reason has to do with off-equilibrium behavior. Suppose, instead that all Neighbors cooperate and play the same local trigger strategies as before. If, say, Neighbor 2 defects to “ D ”, Neighbor 3 must continue to cooperate as prescribed. However, Neighbor 2’s action causes a chain reaction elsewhere: Neighbor 1 then plays “ D ”, then Neighbor 5, until Neighbor 3 faces free riders from both sides. Yet, since Neighbor 3 observed the initial defection, he should have been able to infer Neighbor 4’s reaction 3 periods later. Since LTSE do not allow such inferences, best response behavior may lead individuals away from the set of local trigger strategies. Hence, in some cases, local trigger strategies are inconsistent with the perfection requirement of sequential equilibrium.¹⁶

Consequently, we consider a modification of the perfection requirement in which best responses (off the equilibrium path) may not make full use of relevant information from the past. Call \mathcal{T} an LTSE if for each δ , it comprises a sequential equilibrium *when individuals’ best responses are restricted to be local trigger strategies*. This restriction may be justified when the cost of making out-of-neighborhood inferences outweighs the one period loss from

¹⁶We thank a referee for bringing this point to our attention.

not using the information.

The LTSE defined in this way include the LTSE as defined before. In cliques, the two sets coincide. Previous results therefore remain valid under the new definition.¹⁷ When designs are not cliques, this weakening of the equilibrium concept may admit a greater degree of social cooperation than before. For instance, in the example above, the local trigger strategies in Figure 3 in which all individuals cooperate is now an LTSE here.

5.2 Incentive Constraints

For arbitrary neighborhood designs, LTSE (defined either way) are no longer simply characterized by a single constraint such as Inequality (6). An individual can no longer count on the coordination of punishments of all his neighbors. An additional “perfection” constraint is required. To characterize these constraints, fix an LTSE \mathcal{T} and a realization δ . Let $S = (S_1, \dots, S_m)$ with $S_i \subseteq N_i$ for all i . Define

$$Q^{\mathcal{T}}(S) = \{k \in N : S_k \notin \mathcal{T}_k^\delta \setminus \{\emptyset\}\}$$

so that $Q^{\mathcal{T}}(S)$ denotes the set of all residents $k \in N$ who choose “D” when the set S_k constitutes the neighbors who choose “C”. Given an initial state S^0 , the transition law of motion according to \mathcal{T} is given by $S_i^t = S_i^{t-1} \setminus Q^{\mathcal{T}}(S^{t-1})$ for each $t = 1, 2, \dots$. Here, S_i^t denotes the set of neighbors who choose “C” t periods after initial state S^0 . An *equilibrium path* from S is a sequence $\{S^t\}$ with $S = S^0$ and for $t = 1, 2, \dots$,

$$S^t = (S_1^t, \dots, S_m^t) = (S_1^{t-1} \setminus Q^{\mathcal{T}}(S^{t-1}), \dots, S_m^{t-1} \setminus Q^{\mathcal{T}}(S^{t-1})). \quad (12)$$

If S^* is a fixed point of the transition map defined by (12), then S_i^* is i 's equilibrium set of cooperating neighbors. Observe that $(\emptyset, \dots, \emptyset)$ is a trivial fixed point of this map. The set of cooperators everywhere is the union denoted by: $\mathcal{S}^* = \cup_i S_i^*$.

Since an equilibrium fixed point, S^* , is correctly inferred by all individuals, the transition equation (12) may be used to construct incentive constraints along the equilibrium path. Given S^* with $i \in S_i^*$, let $\{\hat{S}^t\}$ denote an equilibrium path from $S^* \setminus \{i\}$. The latter denotes a continuation following a deviation by resident i from “C” to “D”. Letting $s_i^* = |S_i^*|$ and $\hat{s}_i^t = |\hat{S}_i^t|$, i 's equilibrium incentive to cooperate is given by

$$(s_i^* - 1)c - (n_i - s_i^*)\ell \geq (1 - \delta_i)(s_i^* - 1)d + \delta_i \sum_{t=1}^{\infty} (1 - \delta_i)\delta_i^{t-1}\hat{s}_i^t d \quad (13)$$

Notice that if i is expected to cooperate, then he anticipates $(s_i^* - 1)$ other cooperators in his neighborhood, whereas if he deviates then there are \hat{s}_i^t other cooperators in the

¹⁷If individuals are fully connected, the game is one of perfect monitoring. Hence, grim trigger strategies are best responses to grim trigger strategies.

continuation. In the case of a clique, all of i 's neighbors are able to punish him if he deviates. Then (13) reduces to (6), the simple constraint for cooperation in a clique. Inequality (13) implicitly defines a lower bound, $L_i(S^*)$, for i 's patience parameter δ_i . That is, i chooses “C” in equilibrium if $\delta_i \geq L_i(S)$.

As for the “perfection” constraint, suppose that $S^0 = S_i^* \setminus \{j\}$ is observed by i , and the continuation prescribes that i “punish” j 's deviation. Let $\{S^t\}$ denote the path from S^0 , and let $\{\tilde{S}^t\}$ denote the path from $S^1 \cup \{i\}$. Then the individual's constraint is

$$\sum_{t=1}^{\infty} (1 - \delta_i) \delta_i^{t-1} s_i^t d \geq (1 - \delta_i) [s_i^t c - (n_i - s_i^t) \ell] + \delta_i \sum_{t=1}^{\infty} (1 - \delta_i) \delta_i^{t-1} \tilde{s}_i^t d. \quad (14)$$

Inequality (14) implicitly defines an upper bound, $U_i(S^0)$, for i 's discount factor. If $\delta_i > U_i(S^0)$ then the patient player prefers to tolerate a deviation in his neighborhood for a period. This may occur when, for example, most of his neighbors do not initially observe the deviation. It is not hard to see that $U_i(S^0) = 1$, i.e., the constraint never binds in a clique. Since all individuals in the clique are connected, one can easily construct LTSE so that all individuals punish a deviator.

Inequalities (13) and (14) determine incentive constraints along the equilibrium path or, in the case of (14), after an immediate defection. Generalizations for each may be constructed to define best responses off the equilibrium path as well.¹⁸

5.3 Main Results

What types of designs are optimal when the planner has unlimited choice of design? In this Section we characterize the optimal neighborhood design for two special, but important cases. We first take up the case when the planner has full information. In this case, the distribution G is degenerate, and so the planner can connect particular types, fully anticipating the degree of cooperation resulting from their interaction. In the second case, the types are independently and identically distributed. With this degree of uncertainty, individuals are identical *ex ante*, though *ex post* the planner cannot be certain that particular neighbors will end up compatible.

Curiously, the optimal design in each case does not depend on which definition of LTSE is used. The reason is that while the new LTSE concept admits more social cooperation than before, cooperating individuals are fully connected in the optimal designs in all our main results. Therefore, trigger strategies *are* best responses among those for whom incentives matter.

¹⁸These can be found in an earlier version of this paper: Haag and Lagunoff (1999).

Before proceeding with the aforementioned characterizations, we prove a preliminary result below which reveals why heterogeneity is necessary for the planning problem to be nontrivial. The notation $\text{supp } G$ refers the support of G .

Theorem 2 1. *If $\text{supp } G \subseteq [\frac{d-c}{d}, 1)^m$, then an optimal neighborhood design is uniquely given by the maximal clique: $N_i = M$ for all i .*

2. *If $\text{supp } G \subseteq [0, \frac{d-c}{d})^m$ then any neighborhood design is optimal.*

The result is straightforward, utilizing incentive constraints (13) and (14). The proof is in the Appendix. Part 1, the first support condition, is more important than Part 2. Part 1 covers the typical case of observable, homogeneous discount factors (i.e., distribution G is degenerate and places full mass on the vector $(\hat{\delta}, \dots, \hat{\delta})$ for some $\hat{\delta} \geq \frac{d-c}{d}$). Given the result, the design problem is relatively uninteresting in conditions under which the Folk Theorem applies. Limited interaction is optimal only when irreconcilable social conflicts arise. Part 2 is sensitive to the particular normalization we use in the stage game.¹⁹

Full Information

Suppose that the planner knows the exact value of δ , that is, let G be the degenerate distribution that places probability one on some vector $\delta = (\delta_1, \dots, \delta_m)$. Define the set $\Omega = \{i : \delta_i \geq \frac{d-c}{d}\}$. A preliminary result establishes that in the optimal equilibrium in the optimal design, the individuals in Ω are those that cooperate.

Lemma 2 *If G is a degenerate distribution that places full mass on some δ , then the optimal LTSE in any optimal design satisfies: $\mathcal{T}_i^\delta \neq \{\emptyset\}$, $\forall i \in \Omega$, and $\mathcal{T}_i^\delta = \{\emptyset\}$, $\forall i \notin \Omega$,*

The proof is in the Appendix. Let $|\Omega| = \omega$. The Lemma states that there are precisely ω cooperators and $m - \omega$ uncooperative individuals. To see how they are all connected in the optimal design, we begin by partitioning the cooperators in Ω by their degree of tolerance for free riders.

¹⁹If the payoff to the pair (D, D) is greater than zero, than connected nodes are generally preferable to singletons.

$$\begin{aligned}
\Omega_1 &= \{i \in M : \delta_i \geq \frac{d-c}{d} + \frac{\ell}{d} \frac{1}{\omega-1}\} \\
&\vdots \\
\Omega_k &= \{i \in M : \delta_i \geq \frac{d-c}{d} + \frac{\ell}{d} \frac{k}{\omega-1}\} \\
&\vdots \\
\Omega_{m-\omega} &= \{i \in M : \delta_i \geq \frac{d-c}{d} + \frac{\ell}{d} \frac{m-\omega}{\omega-1}\}
\end{aligned}$$

Using Inequality (6), the set Ω_k is the subset of cooperators in Ω who are able to tolerate being connected to at most k free riders out of $\omega + k$ total neighbors. Note that $\Omega_k \subset \Omega_{k-1}$, and it may be the case that $\Omega_k = \emptyset$ for some large enough index k . We can now describe the full information optimum.

Theorem 3 *Suppose that G is degenerate. Then every optimal neighborhood design satisfies,*

(i) $i \in N_j, \forall i, j \in \Omega$, and

(ii) there exists an ordering, $i_1, i_2, \dots, i_{m-\omega}$, of the uncooperative individuals, such that for each $k = 1, \dots, m - \omega$, $i_k \in N_j$ iff $j \in \Omega_k$.

Since uncooperative individuals in $M \setminus \Omega$ are indistinguishable, all optimal designs are equivalent up to payoff-irrelevant permutations of uncooperative individuals. According to the result, the optimal design exhibits a spatial pattern with a cooperative “core” and an uncooperative “fringe” connected to the more tolerant elements of the core. Since the planner has full information, he knows where the social frictions lie. While all pairs of cooperative individuals are linked, each uncooperative individual is connected to as many cooperators as will tolerate his free riding. The gain to each uncooperative neighbor net of the loss to the cooperator is $d - \ell$.

As for the proof, Part (i) in the theorem is a straightforward consequence of the Lemma. Since connectivity among cooperators increases social welfare, it is clear from the Lemma that all cooperators in Ω should be connected to each other. Part (ii) follows from the assumption that $d - \ell > 0$. When $d - \ell > 0$, there is a net social gain to connecting a free rider to a cooperator provided that the cooperator is not induced to change his behavior as a result.²⁰ Part (ii) is an iterative application of the equilibrium incentive constraint (6).

²⁰If $d - \ell < 0$, then there is no social gain to connecting free riders, and the optimal structure consists of an isolated clique of cooperative individuals.

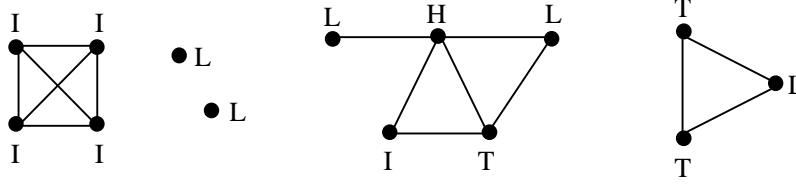


Figure 5:

The degree of slackness in the incentive constraints of cooperators determines the feasibility of adding a marginal free rider.²¹ However, each additional free rider raises the threshold required to tolerate subsequent free riders. Hence, if $\Omega_k = \emptyset$, then at most only $k - 1$ free riders are tolerated in the network. The rest remain unconnected to cooperators.

The graphs in Figure 5 illustrate three examples of optimal designs. The types L , I and T correspond to the familiar three types δ_L , δ_I and δ_T in Section 1. Recall that low types L never cooperate; intolerant types I only cooperate if surrounded by other cooperators; tolerant types T tolerate a single free rider. Type H is assumed to lie in Ω_2 so that this type tolerates two free riders.

Independent Types

The next result establishes cliques as optimal neighborhood designs when types are iid distributed. Since cliques are maximally connected coalitions or communities, one person's decision affects all in his local community. Social cooperation becomes a local public good.

Theorem 4 *If discount factors are distributed iid, then there is an optimal neighborhood design in which M is partitioned into identically sized cliques.*

The idea of the proof, given below in full, is not complicated. Given any design N and any local trigger strategy equilibrium, the planner's criterion $W(N)$ may be decomposed into a weighted sum $\sum_i \frac{n_i-1}{m} W_i(N)$ where $W_i(N)$ is the average value of all the links that include individual i . The importance of this decomposition is that the scale, $\frac{n_i-1}{m}$, of neighborhood size is separated from the individual, average link payoffs $W_i(N)$. That is, each $W_i(N)$ depends on N only through the equilibrium partition $\mathcal{F}(N)$ that describes the measurable events in type space $[0, 1]^m$. When types are iid, $\mathcal{F}(N)$ consists of rectangles: $\mathcal{F}(N) = \times_i \mathcal{F}_i(N)$. In a clique of size n , for example, $\mathcal{F}(N)$ consists of the $\times_i H_{n q_i}$ rectangles. More

²¹This aspect is reminiscent of Bernheim and Whinston's (1990) study of collusion in multi-market firms. The difference is that their design decision is decentralized and limited to one firm's connections.

generally, we show that optimal partitions, i.e., those that maximize $W_i(N)$, must consist precisely of these rectangles $\times_i H_{n_{q_i}}$. These rectangles are generated when incentives to cooperate are maximal, i.e., when a cooperator presumes that all his neighbors will punish his deviation from cooperation. All cliques satisfy this requirement. Finally, since each $\mathcal{F}_i(N)$ enters the optimization problem identically and independently of other $\mathcal{F}_j(N)$, there is an optimal partition which is symmetric, i.e., $\mathcal{F}_i(N) = \mathcal{F}_j(N)$. Symmetric partitions must come from symmetric graphs such as cliques. The optimally sized clique is then shown to have an optimal partition. To sum up, the iid assumption admits optimal symmetric partitions (symmetric graphs), and cliques are symmetric graphs which provide maximal incentives to cooperate.²²

Proof of Theorem 4 Let $([0, 1]^m, \mathcal{B}, G)$ denote the probability space of profiles of discount factors. Here, \mathcal{B} denotes the m -dimensional Borel σ -algebra. Recall the notation $S = (S_1, \dots, S_m)$ in N with $s_i = |S_i|$ for each i and $\mathcal{S} = \cup_i S_i$. It will prove more convenient to work with the set of cooperators \mathcal{S} rather than the vector S of cooperators in each neighborhood.²³ Fix an arbitrary neighborhood design N . We denote the equilibrium mapping from δ to a set \mathcal{S} of equilibrium cooperators by $\mathcal{T}^\delta \mapsto \mathcal{S}$. Define

$$E(\mathcal{S}) = \{\delta : \mathcal{T}^\delta \mapsto \mathcal{S}\}$$

The set $E(\mathcal{S})$ is the “event” in discount factor space in which the set \mathcal{S} is the number of cooperators in the optimal LTSE. In a LTSE, each individual is required to satisfy some incentive constraint of the form $\delta_i \in [a_i, b_i)$. Hence, the event $E(\mathcal{S})$ constitutes a finite union of disjoint rectangles, and so $E(\mathcal{S})$ is a \mathcal{B} -measurable set. By taking intersections of the sets $E(\mathcal{S})$, $\mathcal{S} \in 2^M$, we generate a partition \mathcal{F} of $[0, 1]^m$ into rectangles. When necessary, we write $\mathcal{F}(N)$ to express the dependence of the partition on the neighborhood design N (given the optimal LTSE, \mathcal{T}). We are now ready to define an $\mathcal{F}(N)$ -measurable probability measure P by

$$P(\mathcal{S}|N) \equiv \int_{\delta \in E(\mathcal{S})} dG = \sum_{\{E \in \mathcal{F}(N) : E \subseteq E(\mathcal{S})\}} \int_E dG$$

Since the δ_i are independently distributed, G admits a product measure $G_1 \times \dots \times G_m$. This means that $\mathcal{F}(N) = \times_i \mathcal{F}_i(N)$, and for each rectangle E , $\int_E dG = \prod_{i \in M} \int_{E^i} dG_i$ where each E^i is a $\mathcal{F}_i(N)$ -measurable set. Hence,

$$P(\mathcal{S}|N) = \sum_{\{E \in \mathcal{F}(N) : E \subseteq E(\mathcal{S})\}} \prod_{i \in M} \int_{E^i} dG_i$$

²²While the argument does not rule out other graphs as possibly optimal, we think it is unlikely to be true.

²³From \mathcal{S} , one can easily recover S by setting $S_i = N_i \cap \mathcal{S}$ for all i .

Fix an individual i . Observe that if \mathcal{S} occurs then the total payoff to i if we attribute to him the average payoff of each of his links is $\alpha_{\mathcal{S}}c + (1 - \alpha_{\mathcal{S}})\frac{(d-\ell)}{2}$ if $i \in \mathcal{S}$ and equals $\alpha_{\mathcal{S}}\frac{(d-\ell)}{2}$ if $i \notin \mathcal{S}$, where $\alpha_{\mathcal{S}}$ is the fraction of cooperating neighbors. This can be expressed succinctly as

$$V_i(\alpha_{\mathcal{S}}, I_{\mathcal{S}}) \equiv \left[I_{\mathcal{S}} \left(\alpha_i \left(c - \frac{(d-\ell)}{2} \right) + \frac{d-\ell}{2} \right) + (1 - I_{\mathcal{S}}) \alpha_i \frac{d-\ell}{2} \right]$$

where $I_{\mathcal{S}}$ is the indicator function giving a value of 1 if the event $\{i \in \mathcal{S}\}$ holds, and giving value 0 otherwise. Note that V_i is increasing in $\alpha_{\mathcal{S}}$ and in $I_{\mathcal{S}}$

Now observe that $W(N) = \sum_i \frac{n_i-1}{m} W_i(N)$ where

$$W_i(N) = \sum_{\mathcal{S}} P(\mathcal{S}|N) V_i(\alpha_{\mathcal{S}}, I_{\mathcal{S}}) \quad (15)$$

Note that $W_i(N)$ depends on the design through the partition $\mathcal{F}(N)$. Indeed, it will sometimes prove helpful to express this dependence explicitly by writing $P(\mathcal{S}|\mathcal{F}(N))$. Consider an alternative partition $\mathcal{F}^*(N) = \times_i \mathcal{F}_i^*(N)$ constructed from the design N as follows: for each i let $\mathcal{F}_i^* = \{H_{n_i, q}, q = 0, 1, \dots, n_i - 1\}$. Then let $W_i^*(N)$ denote the planner's payoff constructed from probability $P(\cdot|\mathcal{F}^*(N))$.

Lemma 3 *For each i , $W_i^*(N) \geq W_i(N)$*

The idea of the proof is that W_i^* is to show that each individual is more likely to cooperate when P is measurable with respect to $\mathcal{F}^*(N)$ rather than with respect to $\mathcal{F}(N)$. The proof of this and all subsequent Lemmata are in the Appendix.

Recall that that a choice of design N to maximize $W_i^*(N)$ amounts to choosing a partition $\mathcal{F}^*(N)$. We call a partition that maximizes $W_i^*(N)$ an *optimal partition for i* . Call a partition $\mathcal{F}^*(N)$ *symmetric for N_i* if $\mathcal{F}_k^*(N) = \mathcal{F}_j^*(N)$ for all $j, k \in N_i$.

Lemma 4 *For each i , there exists a symmetric partition for N_i which is optimal for i .*

Lemma 5 *If N is a clique then $\mathcal{F}^*(N) = \mathcal{F}(N)$.*

Let $\mathcal{F}^*(N)$ be a symmetric for N_i and optimal for i . Notice that any other symmetric partition $\mathcal{F}^*(N')$ with $n_j = n'_j$ for all $j \in N_i$ is also optimal for i . A special case of this is a clique. Therefore, without loss of generality, we restrict attention to $\mathcal{F}^*(N)$ in which N is a clique. Then, by the previous Lemma, the actual partition $\mathcal{F}(N)$ induced by the equilibrium sets $E(\mathcal{S})$ is optimal for i . Let N^* denote a clique that maximizes $\frac{n-1}{m} W_i(N)$. The previous Lemmata establish that $\frac{n^*-1}{m} W_i(N^*) \geq \frac{n'-1}{m} W_i(N')$ for each i and each design N' . We conclude that a design N that partitions the population into cliques N^* is optimal. \diamond

5.4 Correlated Types

When the distribution G admits correlation, then incomplete graphs with low neighbor overlap may be optimal for certain regions of the parameter set. To see this, we work out an example with four people and four types. As before, a low type, δ_L , is below the minimal threshold for cooperation and so will always defect. An “intolerant” type, δ_I , will cooperate so long as every one of his neighbors cooperates and defect if any of his neighbors defects. The “tolerant” type, δ_T , is associated with residents who can withstand one defector for every cooperator, but not more.

Consider a distribution G which puts full mass on the permutations from the set $\{\delta_L, \delta_I, \delta_T, \delta_T\}$. The planner knows there are two tolerant types, one intolerant type, and one uncooperative type (but does not know who is which). The payoffs associated with each of the structures are listed under Case 1 of Table 1 below.










Graph Number	Neighborhood Design	Support $\{\delta_L, \delta_I, \delta_T, \delta_T\}$
1		$\frac{1}{2}c$
2		$\frac{1}{2}c + \frac{1}{6}(d - \ell)$
3		$\frac{2}{3}c + \frac{1}{6}(d - \ell)$
4		$\frac{7}{12}c + \frac{5}{24}(d - \ell)$
5		$\frac{1}{2}c + \frac{1}{8}(d - \ell)$
6		$\frac{2}{3}c + \frac{1}{2}(d - \ell)$
7		$\frac{2}{3}c + \frac{1}{3}(d - \ell)$
8		$\frac{1}{4}c + \frac{1}{12}(d - \ell)$
9		0

Table 1

To calculate a particular example, consider graph number 4 in Table 1. With probability $\frac{1}{2}$, the low type lies at one of the end points, in which case there is a $\frac{2}{3}$ probability that the low

type's neighbor is a tolerant type. This gives an average payoff of $\frac{1}{4}(c + c + 2c + (d - \ell)) = (c + \frac{1}{4}(d - \ell))$. There is also a $\frac{1}{2}$ probability that the low type lies at one of the end points and a $\frac{1}{3}$ chance that the low type's neighbor is an intolerant type, yielding an average payoff of $\frac{1}{4}(c + c + (d - \ell)) = (\frac{1}{2}c + \frac{1}{4}(d - \ell))$. Finally, there is a $\frac{1}{2}$ probability that the low type is on the interior with a $\frac{2}{3}$ probability that its interior neighbor is a tolerant type. This gives an average payoff of $\frac{1}{4}(c + c + (d - \ell)) = (\frac{1}{2}c + \frac{1}{4}(d - \ell))$. All other possibilities lead to no cooperation, hence a payoff of zero. Overall, the expected average payoff is $\frac{1}{2}\frac{2}{3}(c + \frac{1}{4}(d - \ell)) + \frac{1}{2}\frac{1}{3}(\frac{1}{2}c + \frac{1}{4}(d - \ell)) + \frac{1}{2}\frac{2}{3}(\frac{1}{2} + \frac{1}{4}(d - \ell)) + \frac{1}{2}\frac{1}{3}0 = \frac{7}{12}c + \frac{5}{24}(d - \ell)$.

Note that in the clique (Graph 9), a “cascade” effect prevents any cooperation. The low type is, by definition, uncooperative which induces the intolerant type to be uncooperative which induces, in turn, the remaining two types to be uncooperative as well, so that no cooperation is possible. Instead, it turns out that the square or closed grid structure (Graph 6) is optimal. The grid balances the connectivity gains when cooperators are linked against connectivity losses when social conflicts, say between types L and I arise.

6 Literature and Extensions

A large literature models the determinants of group size, the effects of congestion, peer effects, and the quality and quantity of public services in local jurisdictions. A small sample includes de Bartolome (1990), Benabou (1993), Conley and Wooders (1998), Epple and Romano (1995), Glomm and Lagunoff (1998, 1999), Oates and Schwab (1991), and Scotchmer (1985) to name only a few.²⁴

The structure of residential developments is a common and interesting local interaction design problem. Social norms of cooperation differ markedly across different communities. In many neighborhoods, for example, cooperative arrangements in supervision of children, in maintaining communal spaces, or in monitoring the safety of the neighborhood are common. In others less so. Typically, cooperative arrangements, when they exist, are not coerced but instead rely on reciprocity and voluntary “good will” of the residents. Detailed discussion, examples, and surveys of the effect of spatial structure on these norms may be found in Logan and Molotch (1987), Landon (1994), Southworth and Ben-Joseph (1997), and White (1980). Landon, in particular, summarized the relative costs and benefits of cul-de-sacs compared to traditional street grids this way:

“In especially successful cul-de-sacs, families get to know one another well. Some cul-de-sacs encourage the development of neighborhood events or stimulate daily social routines... [Yet], on a straight block of houses, people may see what the neighbors across the street are doing, but people do not feel so tightly entwined

²⁴See Conley and Wooders (1998) for further references.

with one another... The cul-de-sac often seems to promote more contact than is desired..." (pp 42-43)

The design of team projects in firms is another application.²⁵ Individual workers differ in their abilities to cooperate, and the project manager may not know these abilities ex ante. The manager would thus find it optimal to maximize interaction between members, but limit the size of teams based on expectations for cooperation.

The present work proposes a tractable way of determining when social contact is likely to be desirable. Problems with local externalities are cast here as mechanism design problems in repeated games. While we are not aware of other work that does this, the nature of the linkage relates the present paper to a growing literature on network externalities. These include a wide variety of applications ranging from macroeconomic growth, e.g., Durlauf (1993), and buyer-supplier relationships, e.g., Kranton and Minehart (1996, 1997), to information transmission, e.g., Bala and Goyal (1998), and social networks, e.g., Chwe (1998).²⁶ Of particular relevance for the present paper is a subset of this literature which examines strategic (noncooperative) behavior in networks.²⁷ Often referred to as *local interaction models*, examples include Anderlini and Ianni (1995), Bala and Goyal (1999), Blume (1993, 1995), Jackson and Wolinsky (1996), Lagunoff and Schreft (1998, 1999), and Morris (1997), among others.²⁸ Analysis of local interaction applied specifically to Prisoner's Dilemma include Epstein (1998) and Tieman, Houba, and van der Laan (1998).

With a few exceptions, these models tend to be either static or assume adaptive adjustment dynamics. By contrast, the present paper studies repeated game effects in graphs with forward-looking agents. In this dimension, a closer analogue may be found in the study of collusion with multi-market firms (see Bernheim and Whinston (1990)), and in the study of multilateral tariff cooperation (see Bagwell and Staiger (1999) and sources contained therein). Also more closely related are general, repeated game models with private monitoring (see, for example, Ben-Porath and Kahneman (1996), Kandori and Matsushima (1998), Mailath and Morris (1999), and Ely and Valimaki (1999)).²⁹ Repeated game effects have also been studied in population games where individuals are repeatedly and randomly paired. See Kandori (1991), Ellison (1994), and Okuno-Fujiwara and Postlewaite (1995). In population games, full cooperation is shown to be supported in repeated Prisoner's Dilemma games

²⁵We thank a referee for this suggestion.

²⁶See also Katz and Shapiro (1994), and Sharkey (1993), and references contained therein.

²⁷This is as distinct from *cooperative* game theoretic models of networks, some references for which can be found in Sharkey (1993).

²⁸The Morris paper is a good source for further references.

²⁹While our model is also one of private monitoring, other papers we have come across tend to either augment the model with communication, or assume approximate perfect/public monitoring. In the latter case, our graph-induced monitoring structure is bounded away from these limiting cases. In the former case, with communication, a Folk Theorem of Ben-Porath and Kahneman (1996) applies to any neighborhood design in which each individual is connected to at least two others.

when discount factors are high enough.³⁰

In addition to local interaction, the present paper introduces heterogeneity in rates of time preference. It is precisely this heterogeneity which makes the linkage design problem nontrivial. Heterogeneity in discounting has been examined by Harrington (1989) who studies the effect of discount rate differences on collusion in oligopolies, and by Lehrer and Pauzner (1999) who examine heterogeneity in general two-player repeated games.

A number of assumptions are used in the analysis for tractability. Four modifications in future work would enrich the present analysis. First, the use of space and distance would add a dimension of realism, particularly when the model applies to residential neighborhood interaction. If, for example, externalities diminish with distance, then a planner may prefer to mitigate the consequences of certain neighbor's actions by increasing space between him and others, rather than exclude the neighbor altogether.

Second, the perfect link between information flows and externalities could be severed. The more interesting case occurs when the externalities flow beyond one's observations.³¹ In such a case, the moral hazard problem is more extreme in larger networks. This suggests a smaller scale of linkage than before will be preferred by the planner.

Third, the design problem could be examined assuming types are private information to the participants themselves. The addition of this type of incomplete information to the imperfect monitoring environment considerably complicates the setup. The current model seems reasonable when types are associated with observed characteristics such as social class or ethnic background.

Fourth, the optimal design problem should be re-examined when nonstationary sequential equilibria are considered. Local trigger strategies entail loss of generality since occasional cooperation is not admitted. The incentive constraints can be relaxed if, say, a relatively impatient type need only choose "C" every n th period.

Finally, a more thorough characterization of the general correlated case is necessary. Although we demonstrate cases where grids are optimal, general conditions under which this is true have not been fully explored.

³⁰Kandori proves a Folk Theorem property for certain Prisoner's Dilemma stage games. Ellison extends the Kandori result to all PD games when a public randomizing device is available.

³¹The case where a player's information extends beyond his externalities is less interesting. To take an extreme case, if residents could condition punishment on the behavior of everyone, then standard repeated game arguments can be used to prove a Folk Theorem when the δ realizations are large enough. Kandori (1992) makes this observation in a random matching model.

7 Appendix

Proof of Lemma 1: As with the standard binomial with independent draws, there are $\binom{n}{s}$ ways to select a set \mathcal{S} with $|\mathcal{S}| = s$ cooperators. Fix one such set \mathcal{S} . Each $i \in \mathcal{S}$ will cooperate if δ_i satisfies Inequality 6 or, equivalently, if $\delta_i \in \cup_{q=0}^{s-1} H_{nq}$. Using the definition of G_{ns} , the probability that each $i \in \mathcal{S}$ cooperates is therefore $(1 - G_{n(s-1)})^s$. Now order individuals in $N \setminus \mathcal{S}$, $j = 1, 2, \dots, n - s$. Utilizing the definition of H_{nq} , consider the first individual $j = 1$ who conditions only on the behavior of members of \mathcal{S} , and does not condition on the other members of $N \setminus \mathcal{S}$. This individual will choose “D” if $\delta_1 \in \cup_{r_1=0}^{n-s-1} H_{n(s+r_1)}$. Suppose that r_1 takes its minimal value $r_1 = 0$. Then this individual could conceivably cooperate if all other members of $N \setminus \mathcal{S}$ cooperate. Conditioning on this possibility, individual $j = 2$ will choose “D” if $\delta_2 \in \cup_{r_2=1}^{n-s-1} H_{n(s+r_2)}$. Note that $r_2 \neq 0$ since he must account for 1’s behavior. Proceeding in this way, if for $j = 1, 2, \dots, k - 1$, r_j takes on its minimal value, then individual k will choose “D” if $\delta_k \in \cup_{r_k=k-1}^{n-s-1} H_{n(s+r_k)}$. The vector $r = (r_1, \dots, r_{n-s})$ we have constructed has the form $r = (0, 1, 2, \dots, n - s)$, and is therefore minimally $n - s$ -admissible. By considering all permutations of the set $N \setminus \mathcal{S}$, and all $r' \geq r$, we obtain the $n - s$ -admissible set, $\mathcal{A}(n - s)$. The probability that all individuals in $N \setminus \mathcal{S}$ choose “D” is then

$$\sum_{r \in \mathcal{A}(n-s)} \prod_{j=1}^{n-s} g(H_{n(s+r_j)})$$

◇

Proof of Theorem 2 (1) Let $\text{supp } G \subseteq [\frac{d-c}{d}, 1)^m$. Then we show that in any clique N' with $m' = |M'|$, the maximal LTSE entails that everyone play C . Any defection is followed by a grim trigger strategy in which everyone plays D thereafter. If such an equilibrium exists, then $W(N'; f) = (m' - 1)c$. Observe that if $m' = m$, i.e., the clique is the maximal one, then the planner’s criterion is maximal over all possible neighborhood designs. We now show that this equilibrium exists, and so the planner can realize $W(N'; f) = (m' - 1)c$. Recall that Inequality (13) is the relevant incentive constraint when $S_i = S_i^*$. Since the grim trigger is played by all in state $S_i^* \setminus \{i\}$, the incentive constraint in (13) reduces to the Inequality in (6). But in the maximal clique, $s_i^* = s_j^* = m$ for all i and j . Therefore, the incentive constraint is $\delta_i \geq \frac{d-c}{d}$ which clearly holds when $\text{supp } G \subseteq [\frac{d-c}{d}, 1)^m$. Finally, observe that the perfection constraint, (14), never binds in the grim trigger LTSE in cliques since in that case (14) reduces to $0 \geq (n_i - 1) - \ell$.

(2) Clearly the bound $(d - c)/d$ is the minimal lower bound for δ_i in any equilibrium of any neighborhood design. Hence, if $\text{supp } G \subseteq [0, \frac{d-c}{d})^m$ then in any neighborhood design the only possible LTSE is the one in which all individuals play “D” every period. Therefore the planner is indifferent between all designs. ◇

Proof of Lemma 2: Suppose that $N = (N_1, \dots, N_m)$ is an optimal design. Suppose, by contradiction, that $a_i^t = D$ for some i, t in the optimal LTSE. Because LTSE are stationary,

we must have $a_i^t = D$ for all t . This stationarity, coupled with fact that G is degenerate, allows us to express (3) as

$$W(N) = \frac{1}{m}(1 - \delta) \sum_t \left(\sum_{j \in M \setminus \{i\}} \delta_j^t u_j(\tilde{a}_{N_j}^t(f)) + \delta_i^t u_i(\tilde{a}_{N_i}^t(f)) \right)$$

Let \mathcal{S} denote the set of cooperators in equilibrium in design N . Let $K = M \setminus (\mathcal{S} \cup \{i\})$ and define $N' = (N'_1, \dots, N'_m)$ by:

$$N'_j = \begin{cases} N_j \setminus \{i\} & \text{if } j \in K \\ N_j \cup \{i\} & \text{if } j \in \mathcal{S} \\ N_i \setminus K & \text{if } j = i \end{cases}$$

In words, N' connects i to all the current cooperators in M , and removes him from all current uncooperative individuals. We now modify the old LTSE as follows. Let $CH(S)$ denote the *comprehensive hull*, i.e., the collection of sets defined by including all supersets of the set S . By writing $CH(\mathcal{T} \cup S)$ to express the collection defined by taking the comprehensive hull of all the sets in \mathcal{T} when S is added to the collection. Let

$$\mathcal{T}_j^{\delta'} = \begin{cases} CH(\mathcal{T}_j^\delta \cup \{i\}) & \text{if } j \in \mathcal{S} \\ \mathcal{T}_j^\delta & \text{if } j \in K \\ CH(\mathcal{T}_i^\delta \cup \{\mathcal{S}\}) & \text{if } j = i \end{cases}$$

We assert that this is indeed an LTSE and is optimal. First, behavior in K is unchanged. Second, current cooperators in \mathcal{S} continue to cooperate with additional incentives since the same set of punishers is strictly larger as it now includes i . It remains to show that i has an incentive to cooperate. Since he is only connected to those in \mathcal{S} and since $i \in \Omega$, $\delta_i \geq L(\mathcal{S}) = \frac{d-c}{d}$. That is, his incentive constraint holds. To verify the perfection constraints of those who would punish i , observe that if i defects, everyone punishes him. Hence, recalling the notation $\hat{S} = S \setminus \{i\}$, we have $\hat{s}_i^t = 0$ for all t . Therefore, $\delta_j \leq U(\mathcal{S} \setminus \{i\}) = 1$.

To verify that \mathcal{T}' is optimal, observe that individual i cooperates and is connected to all current cooperators in \mathcal{S} . He is also disconnected to all free riders. Hence, in every new link created in the new neighborhood structure N' there is mutual cooperation, while in every

deleted link, there had been none. Since \mathcal{T} was asserted to be optimal in N , the LTSE \mathcal{T}' is optimal in N' .

We now show that $W(N') > W(N)$. To see this, observe first that $u_j(\tilde{a}_{N'_j}^t(f)) = u_j(\tilde{a}_{N_j}^t(f)) \quad \forall j \in K$ since all the defecting individuals who were previously connected to i received a payoff of 0 from interacting with a mutually defective individual. Observe next $u_k(\tilde{a}_{N'_k}^t(f)) = u_k(\tilde{a}_{N_k}^t(f)) + c + \ell \quad \forall k \in \mathcal{S}$. Finally, $u_i(\tilde{a}_{N'_i}^t(f)) = |S_i^t|c$, whereas $u_i(\tilde{a}_{N_i}^t(f)) = |S_i^t|(d)$. Then, $W(N') = W(N) + |\mathcal{S}|(c + c + \ell - d) = W(N) + |\mathcal{S}|(2c - (d - \ell))$. Since $2c > d - \ell$, $W(N') > W(N)$. This contradicts the assumption that N was an optimal design. \diamond

Proof of Lemma 3 Let $E(\delta) = \times_i E_i(\delta_{-i})$ denote a rectangle in $\mathcal{F}(N)$ that contains δ . Let $E^*(\delta)$ denote the corresponding rectangle in $\mathcal{F}^*(N)$. We will proceed to show that if $i \in \mathcal{S}$ then $E_i^*(\delta_{-i}) \supseteq E_i(\delta_{-i})$, while if $i \notin \mathcal{S}$ then $E_i^*(\delta_{-i}) \subseteq E_i(\delta_{-i})$. We construct these sets by examining maximal incentives for cooperation.

First, consider $i \in \mathcal{S}$. Let $\delta \in E_i(\delta_{-i})$. We consider the maximal incentives for cooperation. Namely, we suppose that this individual i who cooperates in equilibrium would be punished by all others in N_i should he defect at any stage. Of course, an LTSE in which this is true may not exist. However, if it does exist, then the perfection incentive constraint never binds for any i who cooperates in equilibrium. In fact, these incentives are precisely the incentives given by Inequality (6) since this constraint is generated by cliques in which all individuals in the neighborhood observe a defection. In turn, Inequality (6) is equivalent to $\delta_i \in \cup_{q=0}^{s_i-1} H_{n_i q}$. Since this describes maximal incentives for i to cooperate, such incentives may not hold in an LTSE that generates interval $E_i(\delta_{-i})$. Therefore, $\delta_i \in \cup_{q=0}^{s_i-1} H_{n_i q}$ if $i \in \mathcal{S}$.

Now suppose $i \notin \mathcal{S}$. As before, let $\delta_i \in E_i(\delta_{-i})$. Recall that the maximal incentives for cooperation by i are that all of i 's cooperating neighbors would punish any observed deviation if i chose "C". Under these conditions, however, the proof of Lemma 1 demonstrates that $\delta_i \in H_{n_i s_i + r_i}$ for some $r_i \geq 0$. Since $E_i(\delta_{-i})$ may entail that i choose "D" without the conditional event that all neighbors punish deviation, it follows that $E_i(\delta_{-i}) \supseteq H_{n_i s_i + r_i}$ for some $r_i \geq 0$. \diamond

Proof of Lemma 4 An optimal partition for i is one that solves

$$\max_{\mathcal{F}} \sum_S P(S|\mathcal{F}) V_i(\alpha_S, I_S) \quad (16)$$

By independence, $\mathcal{F} = \times_i \mathcal{F}_i$ and so (16) is equivalent to

$$\max_{\mathcal{F}_1} \max_{\mathcal{F}_2} \cdots \max_{\mathcal{F}_m} \sum_S P(S|\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_m) V_i(\alpha_S, I_S) \quad (17)$$

Notice that the independence of the δ_j implies that each \mathcal{F}_j may be chosen separately, and, of course, the order of optimization is irrelevant. Since $\mathcal{F}_i^* = \{H_{n_s}, s = 0, \dots, n-1\}$ each

solution $\mathcal{F}_i^*(N)$ depends only on the choice n_j of neighborhood size. Since there are only finitely many such choices, at least one must be optimal.

Observe, finally, that V_i is invariant to permutations of N_i such that leave $\alpha_{\mathcal{S}}$ unchanged. Hence, one solution to (17) entails that $P(\mathcal{S}|N) = P(\mathcal{S}'|N)$ for all \mathcal{S} and \mathcal{S}' with $|S_i| = |S'_i|$. Since, for any pair $j, k \in N_i$ the marginal distributions are identical, $G_j = G_k$, this condition will be satisfied if $\mathcal{F}_k^*(N) = \mathcal{F}_j^*(N)$. \diamond

Proof Lemma 5 By the Lemma 1 proof, $E \in \mathcal{F}(N)$ implies $E \in \mathcal{F}^*(N)$. But since both $\mathcal{F}(N)$ and $\mathcal{F}^*(N)$ are partitions of disjoint sets, there are no other possible sets in $\mathcal{F}^*(N)$. Hence $\mathcal{F}(N) = \mathcal{F}^*(N)$. \diamond

References

- [1] Anderlini, L. and A. Ianni (1996), "Path Dependence and Learning from Neighbors," *Games and Economic Behavior*, 13: 141-77.
- [2] Bagwell, K. and R. Staiger (1999), "Regionalism and Multilateral Tariff Co-operation," in *International Trade Policy and the Pacific Rim*, IEA Conference Volume No. 20, J. Piggot and A. Woodland, Eds., Great Britain: MacMillan.
- [3] Bala, V. and S. Goyal (1998), "Learning from Neighbors," *Review of Economic Studies*, 65: 595-621.
- [4] Bala, V. and S. Goyal (1999), "A Non-Cooperative Model of Network Formation," *Econometrica*, forthcoming.
- [5] Benabou, R. (1993), "Working of a City: Location, Education and Production," *Quarterly Journal of Economics*,
- [6] Bernheim, B. D. and M. Whinston (1990), "Multimarket Contact and Collusive Behavior," *Rand Journal of Economics*, 21: 1-26.
- [7] Ben-Porath, E. and M. Kahneman (1996), "Communication in Repeated Games with Private Monitoring," *Journal of Economic Theory*, 70: 281-97.
- [8] Billingsley, P. (1986), *Probability and Measure*, John Wiley and Sons: New York.
- [9] Blume, L. (1993), "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5: 387-424.
- [10] Blume, L. (1995), "Evolutionary Equilibrium with Forward Looking Players," mimeo.

- [11] Compte, O. (1998), "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66: 597-626.
- [12] Conley, J. and M. Wooders (1998), "Tiebout Economies with Differential Genetic Types and Endogenously Chosen Crowding Characteristics," mimeo.
- [13] Chwe, M. (1998), "Communication and Coordination in Social Networks," mimeo, University of Chicago.
- [14] De Bartolome (1990), "Equilibrium and Inefficiency in a Community Model with Peer Group Effects," *Journal of Political Economy*, 98: 110-33.
- [15] Durlauf, S. (1993), "Nonergodic Economic Growth," *Review of Economic Studies*, 60: 349-66.
- [16] Ellison, G. (1994), "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies*, 61: 567-88.
- [17] Epstein, J. (1998), "Zones of Cooperation in a Demographic Prisoner's Dilemma," *Complexity*, 4: 36-48.
- [18] Epple, D. and R. Romano (1995), "Public School Choice and Finance Policies, Neighborhood Formation, and the Distribution of Educational Benefits, mimeo.
- [19] Ely J. and J. Valimaki (1999), "A Robust Folk Theorem for the Prisoner's Dilemma," mimeo.
- [20] Fudenberg, D. and E. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54: 533-56.
- [21] Glomm, G. and R. Lagunoff (1998), "A Tiebout Theory of Public vs Private Provision of Collective Goods," *Journal of Public Economics*, 68: 91-112.
- [22] Glomm, G. and R. Lagunoff (1999), "A Dynamic Tiebout Theory of Voluntary vs Involuntary Provision of Public Goods," *Review of Economic Studies*, 66: 659-77.
- [23] Harrington, J. (1989), "Collusion Among Asymmetric Firms: The Case of Different Discount Factors," *International Journal of Industrial Organization*, 7: 289-307.
- [24] Kandori, M. and H. Matsushima (1998), "Private Observation, Communication, and Collusion," *Econometrica*, 66: 627-52.
- [25] Haag, M. and R. Lagunoff (1999), "Social Norms, Local Interaction, and Neighborhood Planning," mimeo, July.
- [26] Jackson, M. and A. Wolinsky (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71: 44-74.

- [27] Johnson, N., S. Kotz, and A. Kemp (1992), *Univariate Discrete Distributions*, New York: Wiley and Sons .
- [28] Kandori, M. (1992), "Social Norms and Community Enforcement," *Review of Economic Studies*, 59: 63-80.
- [29] Katz, M. and C. Shapiro (1994), "Systems Competition and Network Effects," *Journal of Economic Perspectives*, 8: 93-115.
- [30] Kranton, R. and D. Minehart (1996), "Buyer-Supplier Networks," mimeo.
- [31] Kranton, R. and D. Minehart (1997), "Link Patterns in Buyer-Supplier Networks: Incentives and Allocation in Graphs," mimeo.
- [32] Lagunoff, R. and S. Schreft (1998a), "A Model of Financial Fragility," mimeo.
- [33] Lagunoff, R. and S. Schreft (1999), "Financial Fragility with Rational and Irrational Exuberance," *Journal of Money, Credit, and Banking*, 31: 531-60.
- [34] Langdon, P. (1994), *A Better Place to Live: Reshaping the American Suburb*, Amherst: University of Massachusetts Press.
- [35] Lehrer, E. and A. Pauzner (1999), "Repeated Games with Differential Time Preferences," *Econometrica*, 67: 393-412.
- [36] Logan, J. and H. Molotch (1987), *Urban Fortunes: The Political Economy of Place*, Berkeley: The University of California Press.
- [37] Mailath, G. and S. Morris (1999), "On the Theory of Repeated Games with Private Monitoring: Notes on a Coordination Perspective," University of Pennsylvania mimeo.
- [38] Morris, S. (1997), "Contagion," mimeo.
- [39] Oates, W. and R. Schwab (1991), "Community Composition and the Provision of Local Public Goods," *Journal of Public Economics*, 44: 217-238.
- [40] Okuno-Fujiwara, M. and A. Postlewaite (1995), "Social Norms in Random Matching Games," *Games and Economic Behavior*, 9: 79-109.
- [41] Scotchmer, S. (1985), "Profit Maximizing Clubs," *Journal of Public Economics*, 27: 25-45.
- [42] Sharkey, B. (1994), "Network Models in Economics," *Handbook of Operations Research and Management Science*.
- [43] Southworth, M. and E. Ben-Joseph (1997), *Streets and the Shaping of Towns and Cities*, New York, McGraw-Hill.

- [44] Tieman, A., H. Houba, and G. van der Laan (1998), “Cooperation in a Multi-Dimensional Local Interaction Model,” mimeo, Tinbergen Institute.
- [45] White, M. (1980), *American Neighborhoods and Residential Differentiation*, New York, Russell Sage Foundation.