Social Norms, Local Interaction, and Neighborhood Planning

Matthew Haag[*] and Roger Lagunoff[†]

Revised and Final Version: June 28, 2004[‡]

Abstract

This paper examines optimal social linkage when each individual's repeated interaction with each of his neighbors creates spillovers. Each individuals' 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

[*] The University of Warwick. E-mail: Matthew.Haag@warwick.ac.uk

[†] Georgetown University. E-mail: lagunofr@georgetown.edu.

[‡] This research is partially supported by NSF Grant, SES #0108932 and the Leverhulme Foundation. We also 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 Society for Economic Dynamics, the Western Economic Association, and the Public Choice Society Meetings, the 1999 Summer Econometric Society Meetings, Drew Fudenberg, Nicola Persico, and three anonymous referees of this journal for helpful comments and suggestions.

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 individuals' 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 the planner knows only the ex ante distribution over individual discount factors, then in some cases the optimal design partitions individuals into maximally connected *cliques* (e.g., cul-de-sacs), while in other cases 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.

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.

[1] See, 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).

[2] Related models are discussed in Section 5.

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.

An individual's willingness to cooperate in this game is determined, in part, by the opportunity cost of his time which, in turn, is determined by his discount factor. We allow that discount factors are heterogeneous. Randomly determined discount factors introduce heterogeneity in the population. The planner makes his choice of design knowing only the joint distribution of discount factors 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. The structure of residential developments is, in fact, 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).

The design of team projects in firms is another application. 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.

To obtain useful results, we limit attention to a particular subset of the sequential equilibria. 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 sequential equilibrium in "trigger strategies" in which each individual chooses to condition his cooperation on the continued cooperation of at least one "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 Prisoner's Dilemma games, an individual's cooperation requires a certain threshold level of patience.

In the design problem, we study optimal local trigger strategy equilibria (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.

Whether this trade-off also holds in the general class of sequential equilibria is an open question. Very likely, some generality is lost in our restricting attention to trigger strategies. Later in the paper we examine some of the limitations quite explicitly.[3] On the other hand, remarkably little is known about forward looking equilibrium behavior in local interaction models.[4] Consequently, the restriction to local trigger strategies is a sensible starting point. Trigger strategies have been much studied in the standard literature, and for good reason: they are simple, easily followed norms which nevertheless require forward-looking sophistication of the participants themselves. In our environment LTSE 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 are not aware of other work that examines the *design* of networks in repeated game settings. A complete analysis of the local interaction setting with heterogeneous discounting is a daunting, but worthwhile, task. This paper is therefore best viewed as a first step in this much larger research agenda on optimal, dynamic social interaction.

The outline of the paper and description of results are as follows. Section 2 gives an example that illustrates the basic tradeoff between connectivity and social conflict. The benchmark model is then described in Section 3. Section 4 studies the case in which all groups of connected individuals are maximally connected. Maximally connected graphs, or

[3] See Section 5.1

[4] See Section 6 for a discussion of the proximate literature.

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. 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.[5] Not surprisingly, when everyone is sufficiently patient, the largest possible clique is uniquely optimal. The problem is more interesting when discounting is heterogeneous and 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 ("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

[5] With some notable exceptions (see Harrington (1989), Fudenberg, Kreps, and Maskin (1990), and Lehrer and Pauzner (1999)), the individuals in repeated game models are assumed homogeneous with respect to rates of time preference.

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*. As a useful baseline case, we study symmetric graphs — those designs for which all neighborhoods are of equal size — and IID type distributions. Symmetric graphs with low connectivity (i.e., those with 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. Consequently, among symmetric graphs with IID types, the optimal design partitions individuals into maximally connected graphs, or *cliques*.

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.

We review the proximate literature in Section 6. Section 7 is an Appendix with details of the proofs.

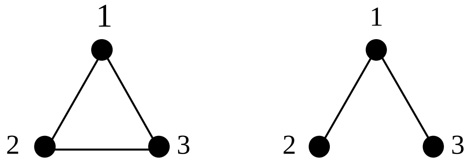


Figure 1: Cul-De-Sac versus Street Grid

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.[6] 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 Prisoner's Dilemma 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, 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

[6] If, for instance, a neighbor enjoys viewing another neighbor's flower bed, then the spillover coincides with his observations of his neighbor.

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	-ℓ, d
D	d,-ℓ	$0, 0$

Figure 2: Prisoner’s Dilemma Game

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 any feasible payoff profile, 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

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

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,

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

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.[7] Instead, he will tolerate uncooperative behavior

[7] If 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 Nth 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,

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., $\frac{d-c}{d} \leqslant \delta_I < \frac{d-c}{d} + \frac{\ell}{d}$.

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 $\delta_1 = \delta_T, \quad \delta_2 = \delta_I, \quad \delta_3 = \delta_L$. Since Neighbor 3 is a low type , full

it may not work when discounting is sufficiently heterogeneous. A referee recently suggested, by way of an interesting example, an asymmetric extension of Ellison’s technique to a local interaction environment. It is hard to tell if the construction is widely applicable, but it does suggest a possible blueprint for incorporating the Ellison technique in some heterogeneous discounting, local interaction settings.

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".

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 $N = (N_1, \ldots, N_m)$. 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 $i \in N_j$. Let $n_i = |N_i|$.

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.[8]

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. 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 $h_i^t =$

[8] See Section 6 for extensions that break this link.

($a^1_{N_i}, a^2_{N_i}, \ldots, a^{t-1}_{N_i}$). 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 $\sum_t (1-\delta_i)\delta_i^t u_i(a^t_{N_i})$ 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 $\delta = (\delta_1, \ldots, \delta_m)$. 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.[9] 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

[9] Only pure strategies are considered in this analysis.

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 $N = (N_1, \ldots, N_m)$, a strategy profile f is anticipated, then the planner's criterion is the expected average discounted value

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

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.

While other welfare criteria could be considered, the utilitarian criterion (3) is a logical starting point for the study of network design because of its implied symmetry. In particular, it treats all links in the graph identically. The utilitarian criterion (3) presumes not only equal treatment of all individuals and all links ex ante, but also weights equally each player's payoff under particular strategy profile f.[10] Note also that by changing payoffs asymmetrically in the game, one can obtain at least some of the optimal designs that could be obtained by varying the welfare weights.

Unfortunately, three issues arise that render criterion (3) inadequate and intractable for

[10] Note that our social welfare criterion presumes that there are no side payments. Otherwise, an optimal strategy profile may entail unbounded borrowing and lending. Also, the normalizations, $(1-\delta_i)$, make more sense if discount factors come from differences in the pure rate of time preference rather than arise from differing frequencies of play across individuals.

our purposes. First, consistency requires that the same type of equilibrium be examined across different types of designs. Otherwise, any design can be shown to maximize (3) provided that the residents play the uncooperative equilibrium following any other chosen design.[11] 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.

Third, even the restriction to equilibria that maximize (3) is, in some sense, too broad. A

[11] Specifically, 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.

recent paper by Lehrer and Pauzner (1999) shows that equilibria can sustain payoffs outside the convex hull of the stage game payoff set if individuals are sufficiently but heterogeneously patient. The intuition is that relatively impatient players can receive favorable payoffs early on, while relatively patient ones receive rewards later in the game. This means that the optimal f in a maximally connected design may be highly nonstationary. Unfortunately, even in games of complete information (i.e., maximal linkage) to date we are not aware of any general characterization of the equilibrium payoff set when discounting is heterogeneous.[12] Hence, for tractability, we introduce in the next Section a restriction to a certain class of stationary equilibria. We would argue that this restriction admits a transparent comparison with uniform discounting models, and, at the same time, highlights potential social conflicts generated by the interaction of patient and impatient individuals.

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.,

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

[12] Lehrer and Pauzner characterize the *feasible* payoff set for 2-player games, and go on to prove a Folk Theorem as discounting approaches one, holding fixed the log ratio between the two players' discount factors.

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.

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,

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

In words, a local trigger strategy in a neighborhood design is one in which an individual agrees to cooperate if and only if i's neighbors who have not yet defected are "acceptable" according to $\mathcal{T}_i^\delta$. Local trigger strategies require each person to bind his behavior to a, perhaps not 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 $N_i(h_i) \notin \mathcal{T}_i^\delta \backslash \{\emptyset\}$, no subsequent cooperating set can be included in $\mathcal{T}_i^\delta \backslash \{\emptyset\}$.

Hereafter, we associate resident i's local trigger strategy with the set $\mathcal{T}_i^\delta$ 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.[13] 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.

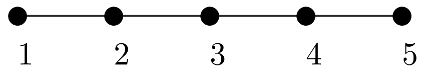


Figure 3: Neighbors on the Line

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

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

"C" every nth period. However, to be successful in an environment with heterogeneous discounting, time varying strategies such as this may require long memory 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.

Lemma 1 *Let $\mathcal{T}$ be a LTSE, and let $\mathcal{S}^t$ denote the set of individuals that choose the cooperative action C in date t in equilibrium. Then $\mathcal{S}^t = \mathcal{S}^\tau$, for all $t, \tau \geq 1$, i.e, LTSE are stationary.*

The proof is straightforward and is contained (along with proofs of all subsequent results) in the Appendix. 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, each neighbor living in a cul-de-sac has full view and interacts with each of the his neighbors, or, each large firm in a public, homogeneous product industry observes the sales of the other firms.

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 symmetric 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

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

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

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

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, equivalently, if

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

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 $H_{n\,q} \subseteq [0,1]$ 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 $H_{n\,n-1} = [0, \frac{d-c}{d})$, and

$$\begin{aligned} H_{n\,q} &= \left[\min\left\{1, \frac{d-c}{d} + \frac{\ell}{d}\frac{[n-1-(q+1)]}{(q+1)}\right\}, \min\left\{1, \frac{d-c}{d} + \frac{\ell}{d}\frac{(n-1-q)}{q}\right\}\right) \\ & \quad if\ q = 0, \ldots, n-2. \end{aligned}$$

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

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

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 $r = (r_1, \ldots, r_k)$ with $r_i \in \{0, 1, \ldots, k-1\}$ for each i is *k-admissible* if there exists a permutation $\theta : \{1, \ldots k\} \rightarrow \{1, \ldots, k\}$ such that the permuted vector r_θ satisfies $r_\theta \geq (0, 1, 2, \ldots, k-1)$. 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, $r = (0, k-1, k-1, \ldots, k-1)$ is admissible. However, $r = (0, 0, k-1, k-1, \ldots, k-1)$ is not.[14]

Theorem 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. The probability derived in Theorem 1 describes a particular type of correlated binomial distribution (“success" likelihoods may be correlated).[15] The purpose of the characterization in the Theorem 1 is to provide an algorithm for computing optimally sized cliques. That is, an *explicit* solution to (5) can be computed using the construction in Theorem 1.

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 2 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. Nev-

[14] Each admissible r can contain at most one “0", two “1"s, three “2"s, and so on.

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

ertheless, one can construct distributions that bound optimal clique size. For example, let α satisfy: $0 < \alpha < G_i(\frac{d-c}{d})$ and $G_i(\frac{d-c}{d} + \frac{\ell}{d}\frac{(1-\alpha)}{\alpha}) = 1$. This last condition implies that no individual will tolerate less than fraction α cooperative neighbors. With this distribution,

$$W(n) < 4c\sum_{s} s\,p(s|n) = 4c\sum_{s\geq\alpha n} s\,p(s|n) \;<\; 4c\sum_{s\geq\alpha n} s\binom{n}{s}\left(1-G_i(\frac{d-c}{d})\right)^{s} G_i\left(\frac{d-c}{d}\right)^{n-s} \tag{7}$$

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 (7) converges to zero as $n \to \infty$. To see this, observe that given the assumptions on α, the right tail, $\sum_{s\geq\alpha n}\binom{n}{s}\left(1-G_i(\frac{d-c}{d})\right)^{s} G_i\left(\frac{d-c}{d}\right)^{n-s}$, is $o(\frac{1}{a^n})$ for some constant $a > 1$. Hence, $\lim_{n\to\infty} W(n) = 0$, and so optimal clique size is bounded.

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. However, before proceeding, we point out a shortcoming in LTSE. For some incompletely connected designs, 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. In this case, Neighbor 3's best response is not a local trigger strategy.

The problem is most evident in designs that are cyclic and for discount factor profiles with highly tolerant individuals. In all the subsequent results, however, these types of "problem" designs do not arise as solutions to the planner's problem. However, an example is Section 6 indicates that cyclical graphs may solve the planner's problem when types are imperfectly correlated. Whether these types of examples are robust, or whether such designs might arise for general sequential equilibria is an interesting, open question.

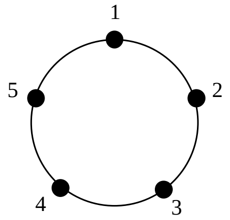


Figure 4: Neighbors on the Circle

5.2 Incentive Constraints

For arbitrary neighborhood designs, LTSE 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. Additional “perfection" constraints are required. We construct two constraints that are necessary conditions for LTSE. Though they are not sufficient conditions, our results show that these constraints suffice to characterize optimal designs in all the environments we examine. Specifically, in these environments, a design that is optimal among all designs sustained by strategies satisfying these constraints is, in fact, sustained by a local trigger strategy equilibrium.

To characterize these constraints, fix an LTSE $\mathcal{T}$ and a realization δ. Let $S = (S_1, \ldots, S_m)$ with $S_i \subseteq N_i$ for all i. Define

$$Q^{\mathcal{T}}(S) = \left\{k \in M : \; S_k \notin \mathcal{T}_k^{\delta} \backslash \{\emptyset\}\right\}$$

so that $Q^{\mathcal{T}}(S)$ denotes the set of all residents $k \in M$ 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 $S_i^t = S_i^{t-1} \backslash Q^{\mathcal{T}}(S^{t-1})$ for each $t = 1, 2, \ldots$. 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 $t = 1, 2, \ldots$,

$$S^t = (S_1^t, \ldots, S_m^t) = (\ S_1^{t-1} \backslash Q^{\mathcal{T}}(S^{t-1}), \ldots, S_m^{t-1} \backslash Q^{\mathcal{T}}(S^{t-1})\ ). \tag{8}$$

In a LTSE, if S_i^* is i's equilibrium set of cooperating neighbors, then $S^* = (S_i^*)_{i \in M}$ is a fixed point of the transition map defined by (8), then Observe that $(\emptyset, \ldots, \emptyset)$ is a trivial fixed point of this map. The aggregate set of cooperators is the union denoted by: $\mathcal{S}^* = \cup_i S_i^*$.

The transition equation (8) can 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 $S^* \backslash \{i\}$. 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

$$(s_i^* - 1)c \ - \ (n_i - s_i^*)\ell \ \geq \ (1 - \delta_i)(s_i^* - 1)d \ + \ \delta_i \sum_{t=1}^{\infty} (1 - \delta_i)\delta_i^{t-1} \hat{s}_i^t d \tag{9}$$

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 continuation. In the case of a clique, all of i's neighbors are able to punish him if he deviates. Then (9) reduces to (6), the simple constraint for cooperation in a clique. Inequality (9) implicitly defines a lower bound, $L_i(S^*)$, for i's patience parameter δ_i. That is, i chooses "C" in equilibrium if $\delta_i \geq L_i(S)$.

A second, "perfection", constraint may be derived as follows. Suppose that $S^0 = S_i^* \backslash \{j\}$ 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

$$\sum_{t=1}^{\infty}(1-\delta_i)\delta_i^{t-1}s_i^t\ d\ \geq (1-\delta_i)[(s_i^1-1)c-(n_i-s_i^1)\ell]+\delta_i\sum_{t=1}^{\infty}(1-\delta_i)\delta_i^{t-1}\tilde{s}_i^t\ d. \qquad (10)$$

Inequality (10) 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 (9) and (10) are necessary conditions for a local trigger strategy profile to be a LTSE. The constraint (9) determines incentives along the equilibrium path, while (10) must hold after an immediate defection. Clearly, the set of local trigger strategies that satisfy (9) and (10) is a superset of the set of LTSE.

Definition 3 A local trigger strategy profile, $\mathcal{T}$, is said to be *expanded optimal* in N if it is optimal in design N among all local trigger strategy profiles that satisfy constraints, (9) and (10). A neighborhood design N is *expanded optimal* if it maximizes the social welfare criterion, (4) , over all designs sustained by expanded optimal trigger strategies.

The notion of expanded optimality will prove useful when characterizing optimal designs in the results below. Roughly, we will show that the resulting expanded optimal designs are sustained not just by the trigger strategies satisfying these incentive constraints, but are sustained by equilibrium (i.e., LTSE) trigger strategies.

5.3 Characterizing Optimal Designs

What types of designs are optimal when the planner can choose from among possibly incomplete designs? 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.

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 *supp* G refers the support of G.

Theorem 2 *1. If 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 supp $G \subseteq [0, \frac{d-c}{d})^m$ *then any neighborhood design is optimal.*

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 $(\hat{\delta}, \ldots, \hat{\delta})$ 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. For if the payoff to the pair (D, D) is greater than zero, then connected nodes are generally preferable to singletons.

The result is straightforward, utilizing incentive constraints (9) and (10). The idea is that, first, we use the constraints to find the expanded optimal design. Next, we show that because the resulting design in Part (i) is a clique, this design can naturally be sustained by a standard, grim trigger strategy which is a special case of a LTSE. Consequently, the design is optimal.

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 $\delta = (\delta_1, ..., \delta_m)$. Define the set $\Omega = \{i : \ \delta_i \geq \frac{d-c}{d}\}$. A preliminary result establishes that in the full information case, all individuals in Ω cooperate in any expanded optimal local trigger strategy.

Lemma 2 *Suppose G is a degenerate distribution that places full mass on some δ. Let N be any expanded optimal design. If $\mathcal{T}$ is a local trigger strategy that sustains N then $\mathcal{T}$ 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.

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

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, N,* *satisfies,*

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

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

Since uncooperative individuals in $M \backslash \Omega$ 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 $d - \ell$.

The neighborhood design described in Theorem 3 characterizes a point on the Pareto frontier which maximizes (4). The welfare weights implied by (4) are identical for all individuals. However, the same design is robust with respect to small perturbations in these welfare weights. To see this, take any pair of individuals. If both are cooperative, or if both are uncooperative, then there is no issue as to whether they should be connected. However, if the link is one in which one individual is cooperative while the other is not, then ideally they should be connected iff $(1-\beta)d - \beta\ell > 0$ where β is a normalized welfare weight between the two. The present welfare criterion assumes that $\beta = 1/2$. Since we assume $d - \ell > 0$, then for small perturbations of weights around 1/2, the currently optimal design remains optimal. If, however, β is enough close to one, then the pair should not be connected.[16] Note that if $d - \ell < 0$, then there is no social gain to connecting this pair unless β is close to zero.

The proof of the Theorem is as follows. First, notice that if the design can be shown to be expanded optimal, then it is, in fact, optimal in the original sense. The reason is that all cooperators form a clique. Since only their incentives need to be checked (uncooperative individuals always play best response by choosing D), there are no perfection constraints. Hence, constraint (9) is both necessary and sufficient to characterize LTSE.

It suffices, therefore, to check that every expanded optimal design satisfies (i) and (ii). 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. Part (ii) is an iterative application of the equilibrium incentive constraint (6). The degree of slackness in the incentive constraints of cooperators determines the feasibility of adding a marginal free rider.[17] 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

[16] We thank a referee for pointing this out.

[17] This 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.

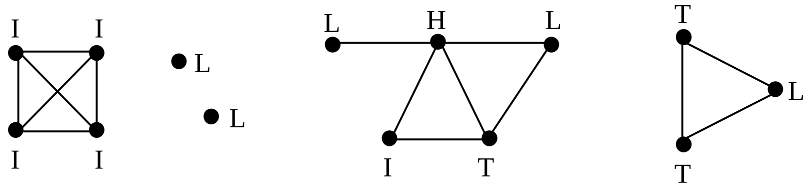


Figure 5: Examples of Optimal Designs

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.

Unknown Types

A design N is *symmetric* if $n_i = n_j$ for all pairs of individuals i and j. Symmetric graphs serve as a useful benchmark for examining the planner's decision problem. With symmetry, analysis is more tractable since each neighborhood is representative, and so a local examination of behavior is without loss of generality. Symmetry may also facilitate commitment by the planner. Consider, for example, a planner who must assign the individuals to the graph when only the ex ante distribution of types is unknown. A planner with egalitarian concerns, for example, has no incentive to re-assign individuals after learning the realized or sample distribution (though not the specific identities) of types.

We show that when the type distribution is iid the optimal neighborhood design in the class of symmetric designs will partition society into cliques of identical size.[18] Since cliques are maximally connected coalitions or communities, one person's decision affects everyone in his local community. Social cooperation becomes a local public good.

Theorem 4 *If discount factors are distributed iid, then for any symmetric design, N, there is a neighborhood design N^* which partitions the population into identically sized cliques, and such that $W(N^*) \geq W(N)$.*

The idea of the proof is quite simple. Cliques are symmetric graphs which are shown to provide maximal incentives to cooperate.

Proof of Theorem 4

Recall that i's action in any period t is denoted by $a_i^t \in \{C, D\}$. It will prove more convenient to denote i's action by its equivalent mixed strategy in $\{0,1\}$. Let $\sigma_i \in \{0,1\}$ denote individual i's action, where $\sigma_i = 1$ is the value of the mixed strategy that places full mass on $a_i = C$ ("cooperate") while $\sigma = 0$ is the value of the mixed strategy that places full mass on $a_i = D$ ("defect").

[18] A referee correctly pointed out that symmetric designs are almost certainly not optimal generally if "remainder problems" are not ignored. For example, it is better in certain cases to leave some of the individuals unconnected if the probability of the uncooperative type is high.

Let $\sigma = (\sigma_1, \ldots, \sigma_m)$ and $\sigma_{N_i} = \sum_{j \in N_i \setminus \{i\}} \sigma_j$.

Fixing, for the moment, an arbitrary (possibly non-symmetric) design N, and given any realized vector δ of discount factors, we define $V_i(\sigma, N_i)$ to be the average payoff of the $n_i - 1$ paired links, each of which include individual i. Specifically, if individual i chooses to be cooperative, i.e., if he chooses $\sigma_i = 1$, then the average of payoffs over all $n_i - 1$ links is $\sigma_{N_i} c + (n_i - 1 - \sigma_{N_i}) \frac{(d-\ell)}{2}$.[19] If, however, individual i chooses to be uncooperative, i.e., if $\sigma_i = 0$, then the average payoff over the $n_i - 1$ links is $\sigma_{N_i} \frac{(d-\ell)}{2}$. Accounting for either of i's choices, this average link payoff can be expressed succinctly as

$$\begin{aligned} V_i(\sigma, N_i) &\equiv \sigma_i \left[\sigma_{N_i}(c - \frac{d-\ell}{2}) + (n_i - 1)\frac{d-\ell}{2} \right] + (1 - \sigma_i)\sigma_{N_i}\frac{d-\ell}{2} \\ &= \sigma_i \left[\sigma_{N_i}(c - \frac{d-\ell}{2}) + ((n_i - 1) - \sigma_{N_i})\frac{d-\ell}{2} \right] + \sigma_{N_i}\frac{d-\ell}{2} \end{aligned} \tag{11}$$

We can now express the aggregate payoff each period as the sum

$$V(\sigma, N) \equiv \frac{1}{m} \sum_{i \in M} V_i(\sigma_i, N_i) \tag{12}$$

Lemma 3 *The function V in (12) is strictly increasing in σ, and, for each i, is an affine function of σ_i.*

The proof of this and all subsequent Lemmatta are contained in the Appendix.

Now, consider any expanded optimal local trigger strategy profile. Along the realized

[19] If, in any of i's links, both individuals cooperate, then the total payoff between them is $2c$. If only one of them cooperates then the total payoff is $d - \ell$, while if neither cooperate, then the payoff is 0.

path of any such profile, actions are stationary. Therefore, let

$$\sigma^*(\delta, N) = (\sigma_1^*(\delta, N), \ldots, \sigma_m^*(\delta, N)\ )$$

denote the action profile along the realized path, i.e., the action profile which maximizes (12) among all local trigger strategies satisfying incentive constraints (9) and (10). Clearly, σ^* depends (explicitly) on both δ and N.

The social planner's expanded optimal criterion can now be described in terms of V:

$$W(N) \ = \int_\delta V(\sigma^*(\delta, N), N)\ dG \tag{13}$$

where, here, $W(N)$ denotes the *expanded optimal*, rather than optimal, planner's payoff given the expanded optimal local trigger strategy.

We now proceed to construct another artificial planner's criterion $\bar{W}(N)$ based on "pseudo-equilibrium" behavior in which each individual believes, possibly falsely, that all other neighbors in his neighborhood N_i will observe and therefore punish a deviant neighbor in N_i. The construction of $\bar{W}$ proceeds as follows.

For each δ and each design N, define a "best response" function $\Phi_i(\cdot\ ; \delta_i, N_i) : \{0,1\}^{m-1} \to \{0,1\}$ by

$$\Phi_i(\sigma_{-i}; \delta_i, N_i) = \begin{cases} 1 \quad if \quad \delta_i \ \geq \ \min\{1, \dfrac{d-c}{d} + \dfrac{\ell}{d}\dfrac{(n_i - 1 - \sigma_{N_i})}{\sigma_{N_i}}\} \\ \\ 0 \quad if \quad otherwise \end{cases} \tag{14}$$

Recalling that Inequality (6) is necessary and, in an optimal LTSE, sufficient condition for i's cooperation in a clique of size n_i, the function $\Phi_i(\sigma_{-i}; \delta_i, N_i)$ takes a value of "1" whenever this same condition holds. Hence, $\Phi_i(\sigma_{-i};\delta_i,N_i)$ may be interpreted as the best response to σ_{-i} if i holds the potentially false belief that all individuals in his neighborhood observe and punish any deviant in N_i.

Let Φ denote the "best response" map for all individuals:

$$\Phi(\sigma;\delta,N) = (\Phi_1(\sigma_{-1};\delta_1,N_1),\ldots,\Phi_m(\sigma_{-m};\delta_m,N_m)\ )$$

Observe that for any pair of individuals, i and j, $\Phi_i = \Phi_j$ up to isomorphism.[20] Moreover, since Φ is increasing in σ, Tarsky's Fixed Point Theorem applies. Hence, there is some σ for which $\sigma = \Phi(\sigma;\delta,N)$. Denoting the explicit dependence on δ and N, let

$$\bar{\sigma}(\delta,N) = (\bar{\sigma}_1(\delta,N),\ldots,\bar{\sigma}_m(\delta,N)\ )$$

be the fixed point which maximizes (12) among all fixed points of $\Phi(\cdot;\delta,N)$. We remark that the vector of actions, $\bar{\sigma}(\delta,N)$, is the natural analogue of the $\sigma^*(\delta,N)$ in the expanded optimal local trigger strategy. The difference is that $\bar{\sigma}(\delta,N)$ is an "equilibrium" of the game in which each individual has an incorrect model of his neighbors' observations about his other neighbors' behavior.

Now define $\bar{W}$ by

$$\bar{W}(N) = \int_\delta V(\bar{\sigma}(\delta,N),N)\ dG \tag{15}$$

[20] Technically, since $i \in N_i$, we cannot define Φ_j on the set $(\sigma_{-i},\delta_i;N_i)$. However, it is clear that by placing individual j in i's position in the graph, his "best response" defined by (14) would prescribe the same behavior given the same discount factor as i.

Lemma 4 *For any design N,*

$$\bar{W}(N) \geq W(N)$$

The idea of the proof is to show that for each δ an individual who cooperates under σ^* will also cooperate under $\bar{\sigma}$. Note that, so far, nothing in the proof has restricted the analysis to symmetric designs.

Lemma 5 *For any symmetric design, N' (where N' satisfies with $n'_1 = \cdots = n'_m = n'$), there exists a neighborhood design N^* which partitions individuals into identically sized cliques of size $n^* = n'$ and which satisfies*

$$\bar{W}(N^*) = \bar{W}(N')$$

Using Lemma 5, for any symmetric design N' there is a partition, N^*, of individuals into identically sized cliques, and $\bar{W}(N^*) = \bar{W}(N')$. Observe that in any clique, there is perfect monitoring within the clique. Therefore, equilibria exist in which each individual can expect that all other individuals observe and punish a deviator. Hence, $W(N^*) = \bar{W}(N^*)$. Consequently,

$$W(N^*) = \bar{W}(N^*) = \bar{W}(N') \geq W(N'),$$

where the last inequality follows directly from Lemma 4. Since $W(N^*)$ is the planner's value of a clique, it can be sustained by the obvious LTSE in which members of each clique play the grim trigger strategy. We conclude the proof. $\diamondsuit\ \diamondsuit$

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.

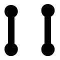
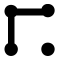
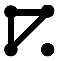
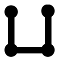
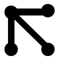
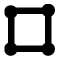
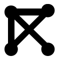
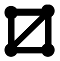
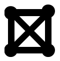
Graph	Neighborhood	Payoffs
Number	Design	
1		$\frac{1}{2}c$
2		$\frac{1}{2}c + \frac{1}{6}\left(d - \ell\right)$
3		$\frac{2}{3}c + \frac{1}{6}\left(d - \ell\right)$
4		$\frac{7}{12}c + \frac{5}{24}\left(d - \ell\right)$
5		$\frac{1}{2}c + \frac{1}{8}\left(d - \ell\right)$
6		$\frac{2}{3}c + \frac{1}{2}\left(d - \ell\right)$
7		$\frac{2}{3}c + \frac{1}{3}\left(d - \ell\right)$
8		$\frac{1}{4}c + \frac{1}{12}\left(d - \ell\right)$
9		0

Table 1: The Permutations with Four Neighbors

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 $\frac{1}{4}\left(c+c+2c+(d-\ell)\right)=\left(c+\frac{1}{4}(d-\ell)\right)$. 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 $\frac{1}{4}\left(c+c+(d-\ell)\right)=\left(\frac{1}{2}c+\frac{1}{4}(d-\ell)\right)$. 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 $\frac{1}{4}\left(c+c+(d-\ell)\right)=\left(\frac{1}{2}c+\frac{1}{4}(d-\ell)\right)$. All other possibilities lead to no cooperation, hence a payoff of zero. Overall, the expected average payoff is $\frac{1}{2}\frac{2}{3}\left(c+\frac{1}{4}(d-\ell)\right)+\frac{1}{2}\frac{1}{3}\left(\frac{1}{2}c+\frac{1}{4}(d-\ell)\right)+\frac{1}{2}\frac{2}{3}\left(\frac{1}{2}+\frac{1}{4}(d-\ell)\right)+\frac{1}{2}\frac{1}{3}0=\frac{7}{12}c+\frac{5}{24}(d-\ell)$.

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. While we cannot say which of the graphs are (fully) optimal, it turns out that the square or closed grid structure (Graph 6) is expanded 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

The present work is a first step toward finding a tractable model of ongoing social interaction. 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 (2000, 2001), to information transmission, e.g., Bala and Goyal (1998, 2001), and social networks, e.g., Chwe (2000).[21] Of particular relevance for the present paper is a subset of this literature which examines strategic (noncooperative) behavior in and/or strategic formation of networks.[22] Examples include Anderlini and Ianni (1995), Bala and Goyal (1998, 2001), Bhaskar (1998) Blume (1993, 1995), Epstein (1998), Houba, Tieman, and van der Laan (2000), Jackson and Wolinsky (1996), Kranton and Minehart (2000, 2001), Lagunoff and Schreft (1999, 2001), Morris (2000), and Vega-Redondo (2003), among others.[23]

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

[21] See also Katz and Shapiro (1994), and Sharkey (1993), and references contained therein.

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

[23] The Morris paper is a good source for further references. Note also: Epstein (1998), Houba, Tieman, and van der Laan (2000), and, Vega-Redondo (2003) all specifically examine local interaction models of Prisoner's Dilemma.

with forward-looking agents. Exceptions are Bhaskar (1998) who studies Prisoner's Dilemma when agents are located on a line, and Lagunoff and Schreft (1999, 2001) who examine the financial fragility of a network, and Vega-Redondo (2003) who examines network formation when payoffs evolve randomly. Close analogues may also 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 (2002)).[24] 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 when discount factors are high enough.[25]

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

[24] While 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.

[25] Kandori 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.

studies the effect of discount rate differences on collusion in oligopolies, by Fudenberg, Kreps, and Maskin (1990) who prove a limited Folk Theorem when some subset of individual's discount factors are zero (perfect impatience), and by Lehrer and Pauzner (1999) who examine heterogeneity in general two-player repeated games.

Finally, our interest in the determinants of social cooperation is not so different from a large literature modeling 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 (2001), Epple and Romano (1995), Glomm and Lagunoff (1998, 1999), Oates and Schwab (1991), and Scotchmer (1985) to name only a few.[26]

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.[27]

[26] See Conley and Wooders (2001) for further references.

[27] The 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

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 nth 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.

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: Fix an LTSE $\mathcal{T}$ and suppose, by contradiction, that it is not stationary. Let $\bar{h}^{\tau}$ denote an equilibrium path history up to (but not including) date $\tau \geq 1$. Let $t > 1$ be the first date at which $\mathcal{S}^t \neq \mathcal{S}^1$. Since, by construction, the set $N_i(\bar{h}_i^t)$ for each i can never increase over time, it follows by definition of local trigger strategies that $\mathcal{S}^t \subset \mathcal{S}^1$. Let $j \in \mathcal{S}^1 \backslash \mathcal{S}^t$. Since $j \in \mathcal{S}^{t-1}$, it follows from the definition of local trigger strategies that $N_j(\bar{h}_j^{t-1}) \in \mathcal{T}_j^{\delta}$ and $N_j(\bar{h}_j^t) \notin \mathcal{T}_j^{\delta}$. But this implies that some other individual at date $t-1$ must have switched from C to D to have induced the change in j's behavior. This contradicts the supposition that t is the first date at which actions have changed. Hence, $\mathcal{S}^t = \mathcal{S}^1$ for all t. ◊

Proof of Theorem 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 $N \backslash \mathcal{S}$, $j = 1, 2, \ldots, n-s$. 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 $N \backslash \mathcal{S}$. 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 $N \backslash \mathcal{S}$ 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 $j = 1, 2, \ldots, k-1$, 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 $r = (r_1, \ldots, r_{n-s})$ we have constructed has the form $r = (0, 1, 2, \ldots, n-s)$, and is therefore minimally $n-s$-admissible. By considering all permutations of the set $N \backslash \mathcal{S}$, and all $r' \geq r$, we obtain the $n-s$-admissible set, $\mathcal{A}(n-s)$. The probability that all individuals in $N \backslash \mathcal{S}$ choose "D" is then

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

$\diamondsuit$

Proof of Theorem 2 (1) Let *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 (9) is the relevant incentive constraint when $S_i = S_i^*$. Since the grim trigger is played by all in state $S_i^* \backslash \{i\}$, the incentive constraint in (9) 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 *supp* $G \subseteq [\frac{d-c}{d}, 1)^m$. Finally, observe that perfection constraints never bind in the grim trigger local trigger strategies in cliques. Hence, (9) is, in fact, a sufficient condition for a local trigger strategy to comprise a LTSE.

(2) Clearly the bound $(d-c)/d$ is the minimal lower bound for δ_i in any equilibrium of any neighborhood design. Hence, if *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. $\diamondsuit$

Proof of Lemma 2: Suppose that $N = (N_1, ..., N_m)$ is an expanded optimal design. Suppose, by contradiction, that $a_i^t = D$ for some $i \in \Omega$ and some t in the expanded optimal trigger strategy. Because local trigger strategies 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

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

Let $\mathcal{S}$ denote the aggregate set of cooperators in equilibrium in design N. Let $K = M\backslash(\mathcal{S} \cup \{i\})$ and define $N' = (N_1', ..., N_m')$ by:

$$N_j' = \begin{cases} N_j\backslash\{i\} \;\; \text{if} \;\; j \in K \\ \\ N_j \cup \{i\} \;\; \text{if} \;\; j \in \mathcal{S} \\ \\ N_i\backslash K \;\; \text{if}\; j = i \end{cases}$$

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

$$\mathcal{T}_j^{\delta\prime} = \begin{cases} CH(\mathcal{T}_j^{\delta} \cup \{i\}) \ \text{ if } j \in \mathcal{S} \\ \\ \mathcal{T}_j^{\delta} \ \text{ if } j \in K \\ \\ CH(\mathcal{T}_i^{\delta} \cup \{\mathcal{S}\}) \ \text{ if } j = i \end{cases}$$

We assert that this is indeed an expanded optimal local trigger strategy profile. 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 (9) holds. To verify the perfection constraint (10) of those who would punish i, observe that if i defects, everyone punishes him. Hence, recalling the notation $\hat{S} = S \backslash \{i\}$, we have $\hat{s}_i^t = 0$ for all t. Therefore, $\delta_j \leqslant U(\mathcal{S} \backslash \{i\}) = 1$.

To verify that $\mathcal{T}'$ is expanded 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 expanded optimal in N, the local trigger strategy $\mathcal{T}'$ is expanded optimal in N'.

We now show that $W(N') > W(N)$. To see this, observe first that $u_j(\widetilde{a}^t_{N'_j}(f)) = u_j(\widetilde{a}^t_{N_j}(f)) \;\; \forall j \in K$ 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 $u_k(\widetilde{a}^t_{N'_k}(f)) = u_k(\widetilde{a}^t_{N_k}(f)) + c + \ell \;\; \forall k \in \mathcal{S}$. Finally, $u_i(\widetilde{a}^t_{N'_i}(f)) = |S^t_i|c$, whereas $u_i(\widetilde{a}^t_{N_i}(f)) = |S^t_i|(d)$. 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 initial assumption that N was an expanded optimal design. ♢

Proof of Lemma 3

Since V is a sum of all the V_i , the requisite properties on V can be verified by showing that they hold on each V_i when restricted to (σ_i, σ_{N_i}).

Using the first right-hand side term in (11), strict monotonicity of V_i in σ_{N_i} is easily verified since $c > \frac{d-\ell}{2} > 0$ was assumed at the outset. To verify strict monotonicity in σ_i, observe that since $\sigma_{N_i} \leq n_i - 1$, and since the second bracketed expression in equation (11) is positive, V_i is strictly increasing in σ_i. Finally, it follows directly from the construction that each V_i is an affine function of σ_i and of σ_{N_i}. ♢

Proof of Lemma 4

Fix some individual i. Recall from the incentive constraints that Inequality (9) is a necessary condition for cooperation in any LTSE. Moreover, if all of i's neighbors are able to punish him if he deviates, then (9) reduces to Inequality (6). Hence, $\sigma_i^*(\delta, N) = 1$ implies that Inequality (6) holds. Using the construction in (14), the Inequality (6) holds whenever $\Phi_i(\sigma_{-i}; \delta_i, N_i) = 1$. Consequently,

$$\sigma_i^*(\delta, N) = 1 \ \Rightarrow \ \text{Inequality (6)} \ \Rightarrow \ \bar{\sigma}_i(\delta, N) = 1$$

Since every individual's cooperation in σ^* implies that same individual's cooperation in $\bar{\sigma}$ for every N and every realization of δ, it follows that $\bar{W}(N) \geq W(N)$. $\diamondsuit$

Proof of Lemma 5

For any N (symmetric or otherwise), let $\alpha_i(N) \equiv \int_\delta \bar{\sigma}_i(\delta, N) dG,$ and $\alpha_{N_i}(N) \equiv \sum_{j \in N_i \setminus \{i\}} \alpha_j(N_j)$. Then let $\beta_{ij}(N) \equiv \int_\delta \bar{\sigma}_i(\delta, N) \bar{\sigma}_j(\delta, N) dG$ and $\beta_{N_i}(N) \equiv \sum_{j \in N_i \setminus \{i\}} \beta_{ij}(N)$. Finally let

$$\alpha(N) \equiv (\alpha_1(N), \ldots, \alpha_m(N)\ )$$

and

$$\beta(N) \equiv (\beta_{N_1}(N), \ldots, \beta_{N_m}(N)\ ).$$

Observe that

$$\begin{aligned}
\bar{W}(N) &= \int_\delta V(\bar{\sigma}(\delta, N), N)\ dG \\
&= \int_\delta V(\bar{\sigma}_1(\delta, N), \ldots, \bar{\sigma}_m(\delta, N), N)\ dG \\
&= \frac{1}{m}\sum_i \int_\delta \Big[\sigma_i(\delta, N)\sigma_{N_i}(\delta, N)(c - \frac{d-\ell}{2}) + \sigma_i(\delta, N)((n_i - 1) - \sigma_{N_i}(\delta, N))\frac{d-\ell}{2} \\
&\qquad\qquad + \ \sigma_{N_i}(\delta, N)\frac{d-\ell}{2}\Big]\ dG \\
&= \frac{1}{m}\sum_i \Big[\beta_{N_i}(N)(c - (d-\ell)\ ) + \alpha_i(N)(n_i - 1)\frac{d-\ell}{2} + \alpha_{N_i}(N)\frac{d-\ell}{2}\Big] \\
&\equiv \bar{V}(\alpha(N), \beta(N), N)
\end{aligned} \tag{16}$$

where the first two equalities follow from the definition of $\bar{W}$ and $\bar{\sigma}$, respectively, the third equality follows from Lemma 3 and equations (11) and (12). The final equality clearly follows from the way that α and β were defined.

Consider any symmetric design N. By the construction of (14), the resulting profile $\bar{\sigma}$ is symmetric, i.e., aggregate cooperation is invariant to permutations of δ. By the iid assumption G admits a product measure $G_1 \times \cdots \times \ G_m$ with $G_i = G_j$ for all i and j, and so $\alpha_i(N) = \alpha_j(N)$ and $\beta_i(N) = \beta_j(N)$ for the symmetric design N. Moreover, the exchangeability property of the iid random variables (δ_i) implies that $\alpha_i(N) = \alpha_i(\hat{N})$ and $\beta_i(N) = \beta_i(\hat{N})$ where $\hat{N}$ is any design derived from N by permutation of the names of the individuals in each neighborhood.

Now, given an arbitrary symmetric N', let N^* be a design that partitions society into identically sized cliques of size $n^* = n'$. Then there is a collection of permutation maps (θ_i) with $\theta_i : N'_i \to N^*_i$ such that $\theta_i(i) = i$ and $\theta_i(j)$ denotes i's neighbor in N^*_i iff j was i's neighbor in N_i'. By our previous argument, $\alpha_i(N^*) = \alpha_i(N')$ and $\beta_i(N^*) = \beta_i(N')$ for each i. Hence, using equation (16), we have

$$\bar{V}(\alpha(N^*), \beta(N^*), N^*) = \bar{V}(\alpha(N'), \beta(N'), N') \tag{17}$$

Using the construction of $\bar{W}$ in (15), equation (17) implies that $\bar{W}(N^*) = \bar{W}(N')$. $\diamondsuit$

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 (2000), "A Noncooperative Model of Network Formation," Econometrica, 68:1181-1229.

[5] Benabou, R. (1993), "Workings of a City: Location, Education, and Production," Quarterly Journal of Economics, 108: 619-52.

[6] Bernheim, B. D. and M. Whinston (1990), “Multi market 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] Bhaskar, V. (1998), “Cooperation and Damage Limitation in a Local Interaction Model," mimeo.

[9] Billingsley, P. (1986), Probability and Measure, John Wiley and Sons: New York.

[10] Blume, L. (1993), “The Statistical Mechanics of Strategic Interaction,” Games and Economic Behavior, 5: 387-424.

[11] Blume, L. (1995), “Evolutionary Equilibrium with Forward Looking Players,” mimeo.

[12] Compte, O. (1998), “Communication in Repeated Games with Imperfect Private Monitoring,” Econometrica, 66: 597-626.

[13] Conley, J. and M. Wooders (2001), "Tiebout Economics with Differential Genetic Types and Endogenously Chosen Crowding Characteristics," Journal of Economic Theory, 98: 261-94.

[14] Chwe, M. (2000), "Communication and Coordination in Social Networks," Review of Economic Studies, 67: 1-16

[15] De Bartolome (1990), “Equilibrium and Inefficiency in a Community Model with Peer Group Effects,” Journal of Political Economy, 98: 110-33.

[16] Durlauf, S. (1993), “Nonergodic Economic Growth,” Review of Economic Studies, 60: 349-66.

[17] Ellison, G. (1994), “Cooperation in the Prisoner’s Dilemma with Anonymous Random Matching,” Review of Economic Studies, 61: 567-88.

[18] Epstein, J. (1998), “Zones of Cooperation in a Demographic Prisoner’s Dilemma," Complexity, 4: 36-48.

[19] Ely, J. and J. Valimaki (2002), "A Robust Folk Theorem for the Prisoner’s Dilemma," Journal of Economic Theory, 102: 84-105

[20] Epple, D. and R. Romano (1995), “Public School Choice and Finance Policies, Neighborhood Formation, and the Distribution of Educational Benefits, mimeo.

[21] Fudenberg, D. and E. Maskin (1986), “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information,” Econometrica, 54: 533-56.

[22] Fudenberg, D., D. Kreps, and E. Maskin (1990), “Repeated Games with Long-run and Short-run Players,” Review of Economic Studies, 57: 555-73.

[23] Glomm, G. and R. Lagunoff (1998), “A Tiebout Theory of Public vs Private Provision of Collective Goods,” Journal of Public Economics, 68: 91-112.

[24] 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.

[25] Harrington, J. (1989), “Collusion Among Asymmetric Firms: The Case of Different Discount Factors, International Journal of Industrial Organization, 7: 289-307.

[26] Kandori, M. and H. Matsushima (1998), “Private Observation, Communication, and Collusion,” Econometrica, 66: 627-52.

[27] Haag, M. and R. Lagunoff (1999), “Social Norms, Local Interaction, and Neighborhood Planning,” mimeo, July.

[28] Jackson, M. and A. Wolinsky (1996), “A Strategic Model of Social and Economic Networks,” Journal of Economic Theory, 71: 44-74.

[29] Johnson, N., S. Kotz, and A. Kemp (1992), Univariate Discrete Distributions, New York: Wiley and Sons .

[30] Kandori, M. (1992), “Social Norms and Community Enforcement,” Review of Economic Studies, 59: 63-80.

[31] Katz, M. and C. Shapiro (1994), “Systems Competition and Network Effects,” Journal of Economic Perspectives, 8: 93-115.

[32] Kranton, R. and D. Minehart (2001), "A Theory of Buyer-Seller Networks," American Economic Review, 91: 485-508.

[33] Kranton, R. and D. Minehart (2000), "Networks versus Vertical Integration," RAND Journal of Economics, 31: 570-601.

[34] Lagunoff, R. and S. Schreft (2001), "A Model of Financial Fragility," Journal of Economic Theory, 99: 220-24 .

[35] Lagunoff, R. and S. Schreft (1999), “Financial Fragility with Rational and Irrational Exuberance,” Journal of Money, Credit, and Banking, 31: 531-60.

[36] Langdon, P. (1994), A Better Place to Live: Reshaping the American Suburb, Amherst: University of Massachusetts Press.

[37] Lehrer, E. and A. Pauzner (1999), “Repeated Games with Differential Time Preferences,” Econometrica, 67: 393-412.

[38] Logan, J. and H. Molotch (1987), Urban Fortunes: The Political Economy of Place, Berkeley: The University of California Press.

[39] Mailath, G. and S. Morris (1999), “On the Theory of Repeated Games with Private Monitoring: Notes on a Coordination Perspective," University of Pennsylvania mimeo.

[40] Morris, S. (2000), "Contagion," Review of Economic Studies, 67: 57-78.

[41] Oates, W. and R. Schwab (1991), “Community Composition and the Provision of Local Public Goods,” Journal of Public Economics, 44: 217-238.

[42] Okuno-Fujiwara, M. and A. Postlewaite (1995), “Social Norms in Random Matching Games,” Games and Economic Behavior, 9: 79-109.

[43] Scotchmer, S. (1985), “Profit Maximizing Clubs,” Journal of Public Economics, 27: 25-45.

[44] Sharkey, B. (1994), “Network Models in Economics,” Handbook of Operations Research and Management Science.

[45] Southworth, M. and E. Ben-Joseph (1997), Streets and the Shaping of Towns and Cities, New York, McGraw-Hill.

[46] Tieman, A., H. Houba and G. van der Laan (2000), "On the Level of Cooperative Behavior in a Local-Interaction Model," Journal of Economics (Zeitschrift fur Nationalekonomie), 71: 1-30.

[47] Vega-Redondo, F. (2003), "Building up Social Capital in a Changing World: A network Approach," Universidad de Alicante, mimeo.

[48] White, M. (1980), American Neighborhoods and Residential Differentiation, New York, Russell Sage Foundation.