The Good, the Bad and the Cunning: How Networks Make or Break Cooperation

Thesis advisor Author

Kenneth A. Shepsle **Jennifer Mary Larson**

The Good, the Bad and the Cunning: How Networks Make or Break Cooperation

Abstract

Groups often find themselves in a position to self-govern: sometimes a formal governing apparatus is weak or nonexistent; sometimes the legal system is underdeveloped, heavily back-logged or inapplicable; and sometimes groups simply have a preference for informal processes. In such cases, contrary to the Hobbesian vision of a self-help nightmare, groups often fare remarkably well both cooperating internally and coexisting with other groups. Diffuse punishment institutions induce cooperation well in tight-knit groups: the theory is well-understood and empirical examples abound. In many realistic settings, though, groups are imperfectly tight-knit, especially when populations are large or sparse or when communications technology is poor (even Facebook networks with very low-cost links are incomplete). Here I relate cooperation to a group's exact structure of communication to identify the role that networks play in making or breaking cooperation. By generalizing the game-theoretic model in Fearon and Laitin (1996), I present a model flexible enough to account for the various ways that a group may be imperfectly tight-knit.

The novel approach presented here makes the study of arbitrary group structures tractable for the first time, which opens the door to comparisons of social structures, of individuals within a social structure, and of potential improvements to a group's cooperative prospects. I argue that the empirical study of networks is important because the exact structure of communication bears on a group's ability to cooperate fully. I show how full cooperation depends on properties of the communication network, and characterize the extent and nature of cheating and lying that become possible when full cooperation is unattainable.

Contents

Acknowledgments

Thanks to the many who have taken time to offer feedback and direction for the project, especially my committee members: Kenneth Shepsle, Robert Bates and James Robinson. I am also grateful for the generous feedback offered by many others, including Bruce Bueno de Mesquita, Andrew Coe, Drew Fudenberg, Michael Gilligan, Torben Iversen, David Lazer, Arthur Spirling and Paul Staniland.

Thanks also to the participants of many seminars at Harvard who listened to iterations of this work as it progressed from prospectus brainstorming to conference presentations to a job talk and, finally, to a completed dissertation.

And of course, a big thanks to my family and friends who made every setback feel smaller and reminded me to enjoy life along the way. Heartfelt thanks go to Andrew Coe for his unwavering support and willingness to spend so many dinners "talking models."

Chapter 1

Introduction

While living on a small island in the Venice Lagoon one summer, despite becoming a regular at the grocery store, I remained only loosely part of the group of Giudeccan women who had been shopping together all their lives. Not once did I observe any of them cutting in front of each other in the line to pay– which would have been social suicide– but *I* was the victim of discrete line-cutting more than once. The offense was safe, since who of consequence to them would I tell? On the flip side, I admittedly sometimes disobeyed the inconvenient rules about the hours trash can be set outside, to the chagrin of my neighbor watching from her window, because running into her or someone she would tell seemed unlikely and so I expected few repercussions from this minor offense. I, as a mere peripheral member of the social group, was a thorn in the side of the cooperative equilibrium on Giudecca, both serving as a tempting target for uncooperative acts, and at times acting uncooperatively myself.

Groups often find themselves in a position to self-govern: sometimes a formal governing apparatus is weak or nonexistent; sometimes the legal system is underdeveloped, heavily back-logged or inapplicable; and sometimes groups simply have a preference for informal processes (see, for example, Colson (1974, ch. 2) or Dixit (2003, ch. 2)). In such cases, contrary to the Hobbesian vision of a self-help nightmare, groups often fare remarkably well both cooperating internally and coexisting with other groups (ch.2 Colson, 1974). The threat generated by diffuse punishment schemes has been found keeping cattle grazing in the right places, fostering trust among traders in underdeveloped markets, and preventing neighboring coethnics from fighting, to name a few (see Ellickson (1991); Landa (1981); Greif (1993); Fearon and Laitin (1996)).

Punishing norm-violators requires knowing who the violators are. Diffuse punishment institutions work well when everyone observes or communicates with everyone else, but in many realistic settings information is imperfect. It is rare that everyone meaningfully communicates with everyone in a group (even low-cost Facebook networks are incomplete), the more so in large or sparse populations or when communications technology is poor. The intuition that dense communication networks are important is old (see Gluckman (1955), Bott (1971)) but theoretical work showing how cooperation depends on communication networks is just catching up.

Existing approaches frame the problem in terms of imperfect monitoring and consider whether private punishment or community punishment institutions can maintain cooperation under various combinations of assumptions (see, for example, Fudenberg, Levine and Maskin (1994), Kandori and Matsushima (1998), Ben-Porath and Kahneman (1996), Dixit (2003), Annen (2011), Lippert and Spagnolo (2011)). The standard approach either omits the exact communication network or restricts attention to a particular class of networks.[1] Here I relate cooperation to a group's exact structure of communication to identify the role that networks play in making or breaking cooperation.

In Chapter 2, "A Failure to Communicate: The Role of Networks in Inter- and Intra-Group Cooperation," I focus on "the good." That is, I explore the conditions under which a group can entice everyone in a group to cooperate. When people are truthful with each other, a group can ensure full cooperation when everyone is talked about enough so that they would face enough punishment by the rest of the community. This happens when links in a communication network are arranged in the right way. Simply having a large number of links and so a large volume of communication spreading around is insufficient to ensure full cooperation. Making sure the links are distributed widely enough so that no one can escape notice is more important than having a lot of links.

With the setup in "A Failure to Communicate," I show that any communication network can be compared in terms of cooperativeness, and any person in a group can be compared based on their position in the network, using a single statistic that pertains to how widely news about people spreads. Group types can then be compared- highly restrictive hierarchies face greater problems cooperating than more egalitarian societies, segregated groups face greater problems than integrated ones, and so on. Groups have options for improving cooperation based on manipulations

[1] There are exceptions– Annen (2011) and Lippert and Spagnolo (2011) take a similar approach to my own and build in a generic network structure. My setup uses a different matching process and adds an out-group. I also use the setup differently, in my case to relate features of the exact structure to cooperation to classify group types, identify structural aspects that matter most, and rank networks with respect to cooperativeness, to name a few.

to their communications technology, punishment code, and social structure.

Chapter 3, "Cheating Because They Can: Networks and Norm Violators," considers "the bad." Whereas "A Failure to Communicate" focuses on fully-cooperative equilibria, this chapter focuses on a class of partially-cooperative equilibria that become possible with certain types of networks. In such equilibria, *most* people are cooperative but some are persistent cheaters, even in the face of punishment. I show that the most likely culprits are those who are not well-integrated into the communication network, like rural fringe populations, or newcomers to a group.

In this chapter, I show that whether cheating is committed against the out-group or fellow in-group members, and whether the peripheral or central members are the most likely targets or victims, depends on the way information spread about offenses. I also take up the observation in Gluckman (1955) that feuds are more likely between people who live far apart. The setup here is ideal for investigating claims like this because it allows the identification of conditions under which we would expect distance to be positively correlated with feuding propensity. In the presence of cheaters, groups face a tradeoff between efficiency and discrimination in designing an institution to restore full cooperation.

Chapter 4, "Deceit, Group Structure and Cooperation," takes on "the cunning." Departing from the assumption of truthful messages in the first two chapters, this chapter accounts for players' incentives to lie to each other. Players gain from punishing others, so would prefer to lie in order to punish more if they could get away with it. The possibility to lie is rampant in an incomplete network (one in which at least one person does not directly communicate with everyone else). I show that even very-but-not-perfectly dense communication networks admit the possibility of lying. To have any hope at detecting lies, a network must allow enough independent verification (by having enough paths between players) and people must remember what they are told and care about it for long enough.

If players punish misdeeds as usual but punish lies more strongly, then groups can maintain cooperation *and* honesty. Inducing honest is easier when peoples' neighbors in the communication network are themselves neighbors to each other. Such "clustering" matters more than degree- having fewer clustered neighbors thwarts lying more effectively than having a greater number of unconnected neighbors. In fact, adding an unconnected neighbor to a person can make that person strictly more likely to lie, despite the greater number of prospective monitors. Interacting with an out-group also encourages more in-group lying. The most profitable lies, though, are not told in order to defect against the out-group more, but in order to frame the in-group and defect against them.

In short, networks matter. How communication spreads from person to person determines how well a group can threaten punishment to keep groups cooperating internally and with other groups. Counting links will not provide a measure of cooperativeness- the arrangement of links is essential and can compensate for sparse communication. Networks with large diameters require especially good communications technology and long grudges to ensure full cooperation. The existence of very isolated group members creates a danger of persistent cheating. Unclustered neighborhoods provide opportunities to lie. People who control information channels can tell lies without detection.

These results are especially useful in light of the influx of networks data and improvements in network elicitation techniques (see, for example, Rao, Mobius and Rosenblat (2007), Conley and Udry (2010), Barr, Ensminger and Johnson (2010)). My hope is that by adding a rigorous theoretical foundation, this work will help to clarify existing empirical work on group structure and guide the design of fruitful future empirical studies.

Chapter 2

A Failure to Communicate: The Role of Networks in Inter- and Intra-Group Cooperation

Groups often need to coordinate their own solutions to uncooperative behavior. We know success is possible for groups trying to share common pool resources (Ostrom, 1990), trade in perilous enviornments (Landa, 1981), extend credit in migrant communities (Laszlo and Santor, 2009), prevent trespassing (Ellickson, 1991), and coexist peacefully with other ethnic groups (Fearon and Laitin, 1996), to name a few. For such groups, decentralized punishment enforces cooperation well so long as everyone knows and communicates with everyone else; news spreads so rapidly that any misdeed can be punished swiftly and thoroughly. When direct communication is *not* ubiquitous, though, some people are more likely to get away with offenses than others, some are more tempting targets than others, and ensuring universal cooperation becomes more difficult. Exactly how difficult depends on who communicates with whom; the communication network makes or breaks the enforcement of cooperation.

I generalize Fearon and Laitin (1996) to flesh out the relationship between a group's communication network and its cooperative prospects. At first glance, it would be easy to dismiss the study of networks in favor of simpler intuitions like: the more communication, the more cooperation. It turns out that this intuition is too simple. A count of communication channels (links) misses the crucial role that the arrangement of channels plays in speeding or slowing the spread of news (and so in promoting or thwarting punishment). I show that groups can have more communication channels but be *less* able to sustain cooperation. It is the network structure as a whole that matters, and the model presented here reduces the multidimensional problem of comparing two networks to a straightforward single-dimensional problem. As a consequence, the model permits the comparison of any two groups with respect to cooperation, any two people in a group with respect to trustworthiness, and any two candidate changes to a group in terms of improved prospects for cooperation.[1] I discuss the most problematic forms of segregation and hierarchies, the benefits of low-diameter social groups, and the threats to full cooperation that come from poorly-integrated group members.

Social scientists are amassing more and more network data from communities across the globe.[2] While a description of a community's communication network is

[1]These comparisons are useful in their own right for understanding and improving group outcomes, but are also particularly useful in light of the growing body empirical work finding that social structure must matter in some way for cooperation. See, for example, Hoff and Fehr (2010), Dunning and Harrison (2010), or Lewis (2010).

[2]See, for example, Ensminger (2001), Banerjee et al. (2011), Conley and Udry (2010), Jackson, Rodríguez Barraquer and Tan (forthcoming).

interesting in its own right, this work pushes forward theory that connects descriptions of networks to group outcomes. A group's communication network transmits information that allows decentralized punishment to keep groups cooperating; some networks are more amenable to cooperation than others. Not only can groups and group members be compared, but the model also produces a single statistic that captures the incentive to act uncooperatively. In light of the debate over which off-the-shelf network statistics are most appropriate in which contexts (see Ensminger (2001) for an overview), a theory-driven statistic with clear meaning is useful. By adding in underlying theory we can move beyond mere description and understand the implications of variance within and across network structures. Knowing the internal structures of groups– their networks– is essential for understanding informal governance and the persistence or breakdown of cooperation.

The Role of Information in Cooperation

Cooperation- why it sometimes persists and when it is likely to break down- is central to topic areas that run the gamut of the social sciences, from interethnic conflict to legislative bargaining to trade. Cooperation can mean obeying social conventions, trading honestly, attending gatherings peacefully, contributing to projects fairly, keeping each other's dangerous secrets, and so on, despite incentives to cut in line, abscond with goods, throw a punch, free-ride off of others' contributions, or rat-out a neighbor. When groups interact with other groups, *inter*-group cooperation is also salient: keeping co-ethnics from provoking interethnic conflict, keeping traders from thwarting intergroup trade, keeping caucus members from undermining bicameral agreements, or keeping gang members from sparking inter-gang warfare.

Empirical work reveals a relationship between groups' ability to cooperate and their social structures. Cross-group social interactions between Hindus and Muslims mitigate interethnic violence in Indian cities (Varshney, 2003). A history of inter-group trade fosters enduring peaceful coexistence among ethnic groups (Jha, 2008). Assignment to a social caste has persistent effects on caste members' responses to non-cooperation, even long after the official eradication of the caste system (Hoff and Fehr, 2010). The structure of kinship and the presence of fictive kinship or laughing cousinage matter for the susceptibility of groups to polarization (Dunning and Harrison, 2010). The onset and success of violent insurgency depends on the organizational structure of insurgent groups (Lewis, 2010). While a relationship appears to be present, thus far we lack a framework useful for understanding disparate results like these and extrapolating to other social structures.

One loosely-defined type of social structure, 'tight-knittedness', is thought to promote cooperation via what Dixit calls "Relation-Based Governance" (Dixit, 2004). Ethnic groups can police their own members to prevent conflict within their group and with other groups (Boehm, 1984; Fearon and Laitin, 1996; Jha, 2009). Experimental evidence suggests that ethnic groups rely on reputation mechanisms and punishment to maintain cooperation, and can do so better than less-cohesive groups (Habyarimana et al., 2009). Communities under minimal governance can stave off self-help and violence with threats of feuds (Colson, 1974, ch. 3). Bands of traders do well to form groups which threaten banishment if trade is conducted unfairly, which can be particularly effective at maintaining efficient levels of trade and minimal theft if implemented among close-knit co-ethnics (Landa, 1981; Greif, 1989; Greif, Milgrom and Weingast, 1994; Rauch and Casella, 1998). Small groups in which monitoring is easy often avoid misuse of common pool resources (Ostrom, 1990). Cattle ranchers who can easily identify each other can prevent trespassing and overgrazing (Ellickson, 1991) and herdsmen made an effort to create close knit groups to employ these kinds of arrangements on the American Western frontier (Anderson and Hill, 2004). All seem to agree that when groups are not perfectly tight-knit, cooperation is more difficult. Despite agreement on the general principle, the exact relationship has not been fleshed out in a way that allows different types of groups to be compared.[3] However, the relationship has not been fleshed out in a way precise enough to permit the study of other social structures– any that are imperfectly tight-knit– or the comparison of classes of structures. Here I aim to do just that.

While there is one way a group can be perfectly tight-knit (everyone knows everyone; the complete network), there are many ways a group can be imperfectly so. Incomplete communication networks– networks in which not all possible links are present so that not everyone directly communicates with everyone else- are probably good descriptions of groups with features like large populations, or time constraints, or high turnover, or limited communications technology, or sparse populations, to name a few. In fact, even networks making use of nearly-costless, nearly-instantaneous technology like Twitter and Facebook are incomplete (Mislove, 2009).[4] Building in

[3]Such informal governance schemes are undertaken not only in weak and failed state contexts but also when legal systems are overworked or when formal legality hasn't been well established or even when such institutions just process claims more quickly or desirably (see, for example, Dixit (2003))

[4]While it is easy to throw around the adage "everyone knows everybody," it is a different thing to mean that all communicate meaningfully, regularly and directly with everyone else. Patterns of

an incomplete network to models of cooperation is more realistic and allows social structures (other than "perfectly tight-knit" groups) to be studied.

The approach here is akin to the study of cooperation in the Law and Economics tradition, in which there exist a few approaches to account for limited information in groups. One approach asks how to maintain cooperation when information is extremely poor. The solution tends to be the installation of a centralized body like a "law merchant"(Milgrom, North et al., 1990) or a mafia (Bandiera, 2003), which might be the lowest-cost option with so little communication (Li, 2003). Here, the interest is in how groups with middling amounts of communication problems– communication isn't perfect, but also isn't zero– fare with decentralized schemes. Existing consideration of middling amounts of communication tend to ignore structure and assume the information problems are experienced uniformly across the population (see, for example, Fafchamps (2002) or Harbord (2006)), or implicitly or explicitly assume a particular form of the communication network (Cooter and Landa (1984),Casella and Rauch (2002), Dixit (2003)). My approach leaves the communication structure general so that any two networks, not just two within a particular class, can be compared.

A second tradition considers the theory of information constraints more abstractly, the most relevant being the study of imperfect monitoring. Imperfect monitoring can mean that players have access to a public signal but noisy signal about the other players (Abreu, Pearce and Stacchetti, 1990; Fudenberg, Levine and Maskin, 1994; Lehrer, 1990), or that players only observe a strict subset of other players ("private

activity between friends on Facebook are even sparser (Wilson et al., 2009).

monitoring"), which is closer to the approach presented here. Fudenberg and Levine (1991) look at imperfect private monitoring without communication, Kandori and Matsushima (1998) look at noisy signals, and Ben-Porath and Kahneman (1996) look at private monitoring with public announcements of the learned information.

The approach here can be understood as private monitoring in which player's private and possibly overlapping information is unveiled to them over time (as news travels through the social network). In contrast with the usual approach, though, the amount of information each player receives can vary based on their position in the social network. This approach has three advantages. One, it leaves the structure of private signals general so that different groups with different communication networks can be compared, and explicitly relates the signals to a social structure. Two, it allows heterogeneity in the amount of information each player receives so that individuals within a group can be compared. Third, the current approach has clear real-world motivation, and produces results applicable to empirical work already underway.[5]

The Model: Building in a Network

I begin with the model of inter- and intra-group interactions in Fearon and Laitin (1996), based on the model of intra-group interactions in Calvert (1995). In the model, two groups– ethnic groups in the original– interact mostly with members of their group but sometimes with members of the other group. Although players

[5]This is in contrast with the standard approach in this tradition in which the particular form of imperfect monitoring under consideration seems to be chosen more for its convenient theoretical properties than for its empirical relevance. When some empirical motivation for a setup is offered, it tends to be accompanied by little discussion of empirical interpretation or relevance of the results, leaving much of this literature to stay within the domain of pure theorists. My hope is that the current model, while less general than the standard private monitoring piece, trades generality for concreteness and will be useful outside the strictly theoretical literature.

recognize every one of their fellow group members, they cannot recognize individual members of the other group. This makes standard reputation strategies more difficult, but Fearon and Laitin (1996) shows that so long as everyone within a group learns who plays whom and the outcomes of all the play in every period, players can police their fellow group members and successfully induce cooperation.[6]

In Fearon and Laitin (1996), the authors acknowledge that assuming everyone in a group is perfectly informed about everyone else in the group might be a stretch, but argue that the omniscience assumption is a useful first approximation for their application since "ethnic groups are frequently marked by highly developed systems of social networks that allow for cheap and rapid transmission of information about individuals and their past histories." This first approximation is useful for showing that *if* everyone learns everything about everyone immediately, then cooperation is possible through this institution. To move forward and answer questions like will this institution succeed if not everyone is informed immediately, how rapid is rapid enough, and which social structures are most cooperative, we must drop the first approximation.

In what follows, I relax the assumption of perfect information within a group and take a closer look at the possible structures that transmit information within a group. Building in a possibly incomplete communication network requires a few additional modifications to the original model. First, the way information spreads must be specified. Second, players' strategies must account for the possibility that they are

[6] In-group policing is one of two efficient subgame perfect equilibria with desirable robustness properties. The other is a spiral equilibrium in which one group indiscriminately punishes all of the other group in the face of intra-group wrongdoing. Here I focus on a group's ability to police its own members.

missing information. And third, players must calculate their expected gains and losses in order to choose their best response in the face of possibly limited information.

Accounting for these three can quickly make the model intractable. What should a player's strategy instruct her to do if she encounters a player who is a defector but she was out of the loop and does not know this yet? If she does not punish, should she be punished even if she didn't know she was supposed to punish? Does a player use information about the network structure to infer who is likely to be cooperative and who is likely to defect? Fortunately, real humans cope with this kind of uncertainty all the time, so we can look to existing punishment conventions for guidance. I make use of a "presumption of innocence" strategy, which instructs players to treat their partners as if they are cooperative unless players have good reason to believe otherwise. Such a convention is codified in many formal punishment regimes, and appears to be at play in informal regimes as well.[7]

The Model: Setup

Begin with two groups, A and B, each containing a set of players $N = \{1, \ldots, n\}$ with n even.[8] Players in the model perform two actions: they play games, and they

[7]Such a convention is sensible in an environment of uncertainty, and punishing confirmed wrongdoing instead of incentives to do wrong makes evolutionary sense since it rewards erring in the direction of cooperativeness.

[8]When n is even, everyone in a group can be matched into a round each period. For simplicity groups are assumed to be the same size. If groups were different sizes, the sizes of the groups must both be odd or both be even for the same to be true. If the total number of players were odd, one player would sit out each period. This would not change the character of the results here– players would all face a probability of not playing in a given period. When groups are the same size n, complete matching is trivially possible since $2n$ is even.

gossip about games. Players play games with one other player at a time, and their opponents are randomly selected each period. With probability p an agent is paired with a member randomly selected from the other group; with probability $1-p$ an agent is paired with a randomly selected member of her own group.[9] Each pair plays a single round of the prisoner's dilemma with common payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,1 & -\beta,\alpha \\ \alpha,-\beta & 0,0 \end{pmatrix} \end{array}$$

where $\alpha, \beta > 1$ and $\frac{\alpha-\beta}{2} < 1$. Players are re-matched every period, and periods occur indefinitely. A player's total payoff is then a stream of discounted single round payoffs. Players have common discount factor $\delta < 1$.

Each period game between players represents a random interaction with the possibility of defecting. Interactions of roughly this sort occur in many arenas: between buyers and sellers in a market who may trade fairly or cheat, between passers-by on a street who may offer pleasantries or sling insults, between people sharing a bar or concert or festival who may tolerate each other's presence or throw a punch. In each instance, the defection is likely to be gossip-worthy: the cheated trader may warn fellow traders. The victim of a fight at a festival will share his anger with friends.

In the model, players gossip about the results of games. Information is transmitted within group A along a communication network (or "gossip network") where the

[9] For simplicity, groups are assumed to be the same size, so this matching is possible. If groups were different sizes, there would be a pair of probabilities such that when some proportion of group A is matched with group B, this implies some different proportion of B is matched with A. The group sizes will place bounds on possible out-group matching probabilities.

nodes are the members of A and the links are communication channels between pairs of players in A, and likewise for information transmitted within B. Formally, a gossip network is a pair (N, g) with a set of nodes N and real-valued $n \times n$ adjacency matrix g such that entry $g_{ij} = 1$ indicates the presence of a link between players i and j, $g_{ij} = 0$ indicates the absence of a link between players i and j. Let $\mathtt{G(N)}$ be the set of all undirected, unweighted, loop-free gossip networks for players in N.[10] The network $g_A(N) \in \mathtt{G(N)}$ is assumed to be common knowledge among players in A, and the network $g_B(N) \in \mathtt{G(N)}$ is assumed to be common knowledge among players in B.[11]

As in Fearon and Laitin (1996), players can perfectly identify members within their own group–they know who they are playing and would recognize them if they are matched again in the future. When randomly paired with a member of the other group, a player only knows he's playing 'someone' from the other group. The group structures are distinct; no links span the two groups. Players in A then at most know the shape of the network in B, but not who sits where.[12] That players cannot recognize individual out-group members may not be a perfect description of some groups, but this scenario is the hard case for intergroup cooperation. Any cross-group links spanning the two groups' communication networks open additional

[10]That is, $\forall i, j \in N$, $i \neq j$, $g_{ij} = g_{ji}$, $g_{ij} \in \{0, 1\}$, and $g_{ii} = 0$.

[11]While the assumption that the network is common knowledge is strong, this setup allows a weaker set of assumptions than in the original. When everyone is assumed to know about everyone as in the original, a complete gossip network is implicitly assumed and must also be common knowledge. However, a particularly strong additional assumption is required about the *content* of the gossip, namely that everyone knows all relevant information about everyone. Here this assumption is weakened.

[12]In fact, players do not need to know anything about the other group's network to play their equilibrium strategy. This assumption, that they know the shape of the other group's network, simply makes the arrival at the equilibrium presented here more plausible. This way all players of group A could assess how well group B can police their own and vice-versa.

punishment options and so make cooperation weakly easier. If punishment can be sustained without any cross-group ties, it can also be sustained with some cross-group ties so that some members in A can recognize some members in B. Of course, the assumption may describe some groups reasonably well. Examples of ethnic groups that have a difficult time recognizing out-group members can be found in Fearon and Laitin (1996, p. 727); even in modern-day Belgium, analysis of cell phone data found that 98% of calls were made within regional groups and only 2% spanned the groups (Blondel, Krings and Thomas, 2010).

At the end of each round, information spreads to neighbors about what happened in the rounds. Such information can be loosely called "gossip," but may spread via talking *or* via observing. Gossip spreads deterministically and expires after a certain number of rounds. The expiration captures the intuition that the memory of misdeeds may not be infinite- once players have carried out punishment for the appropriate duration, the slate is wiped clean (or forgotten). Two features of gossip must be specified: its content and its transmission.

Players gossip about their own rounds and pass along gossip they have heard about the rounds of others until the gossip expires. The content of gossip is a message $\mathbf{m}_{i,t}$ sent from player$i$ to all of i's neighbors at the end of time t. The message contains all unexpired gossip i has heard up until time t plus gossip about is most recent round: who i played and whether either player defected. In other words, if gossip expires after two rounds, the message sent from i at the end of time t might look like: "I just played Bob, who defected against me. Last round, Jill and I cooperated, Ken and Jim cooperated, Andy cooperated with someone from the other group, and Sue was wronged by Liz." In the first pass of the model, the spread and content of the gossip is deterministic. The message about i's round can be understood as information that i communicates to i's neighbors, or as information that i's neighbors observe if i's interaction is public and observable.

In the first pass, then, the content of the messages is not chosen strategically- players have no option to lie about their actions or motivations. This is an admittedly strong assumption, since, if they could get away with it, players would like to select their message to claim that they were defecting out of punishment (and so should incur no punishment themselves) rather than defecting out of malice. There are a number of reasons to proceed with a first pass which assumes honest messages are sent before tackling strategic incentives to lie. The tractability reason is obvious but should not be taken lightly since the introduction of a network opens the floodgates to a deluge of subtleties that must be accounted for even with honest messages.

More compellingly, though, the assumption of truthful gossip may not be so far-fetched. Groups may develop a norm against lying or crying wolf in other contexts and apply that norm generally even if recalculation in the new context would show gains to lying. Other contexts might contain severe consequences to lying: since links in a social network are valuable, trustworthiness is particularly important among linked individuals and deceit risks severing those valuable links (Karlan et al., 2009). An evolutionary view supports the idea that truth-telling may be adopted because of its value in at least some contexts- for example, sociobiology identifies an evolutionary mechanism favoring truthful gossip given its role in fostering reputation mechanisms for cooperation in mobile and dispersed groups (Dunbar, 1998; Enquist and Leimar, 1993). Applying norms from one context to another may be especially likely in groups in which members interact in many different ways- business interactions and every-day interactions and social interactions, say.[13] Ethnic groups have this feature, and tend to be characterized by trust among members (see, for example, Horowitz, 1985). More simply, experimental evidence suggests that when given the opportunity to gossip about others, players in cooperation games do so truthfully (Sommerfeld et al., 2007). We might also imagine that this game is embedded in a larger game not modeled here which contains large enough risk of being discovered and punished for lying that players play this game truthfully. Elsewhere I consider the possibility that none of this reasoning applies, and discuss the options for sustaining cooperation *and* truth-telling when some would act on the strategic incentive to lie (in Chapter 4). For now, the model will take gossip as truthful.

Such is the content of gossip. Gossip spreads from players to their neighbors in the communication network at the end of a round. In some instances, time only allows this single round of gossip before the next round of play occurs. In others, time allows players to pass along the new information, and then pass along the newest information, and then pass along the newest information, and so on, before the next round of play. Call $r = \frac{\textit{Degrees Spread}}{\textit{Rounds Played}}$ the rate at which gossip spreads relative to the rate of play. When $r = 1$, gossip spreads just as fast as games are played: in each time period, one game is played and then gossip spreads one step, from a player to neighbors of that player. When $r > 1$, gossip spreads faster than play so that there is time to pass along new information multiple times between each prisoner's

[13] This is the idea of multiplex networks in Sociology. See, for example, Fischer (1982); McPherson, Smith-Lovin and Cook (2001), p. 437.

dilemma interaction. When this is the case, news reaches more people in time for the next round of play. In-person interactions and email communication is an example in which news may travel faster than play occurs. $r < 1$ means news travels so slowly that multiple interactions may occur before anyone other than the participants learns about the results. This may occur when interactions take place far from home, say at a distant marketplace, so that a few market interactions occur before players have a chance to return home and gossip with their friends and family. The message $\mathbf{m}_{i,t}$ spreads r degrees from player i before the beginning of $t+1$.

Consider the stylized network in Figure 2.1. Here there is a clique with players 8, 7, 6 and 5 all communicating with each other, some players communicating with few others, and player 1 who only communicates with 1 other. Gossip that begins with player 8 reaches four other players immediately: players $9, 7, 6$, and 5. In contrast, gossip that begins with player 1 reaches only one other player immediately, player 2. Figure 2.2 contrasts the reach of gossip in $t+1$ that begins with player 1 in t with gossip that begins with player 8 in t when $r = 1$. Figure 2.3 shows the reach of gossip starting from both sources by $t+2$.

Suppose players 1 and 8 are considering defecting against the out-group (not pictured) in period t. If the rate of gossip is 1 (that is, $r = 1$), then only one other player could possibly punish player 1 in $t+1$ whereas four other players could punish player 8 for his defection. This is illustrated in Figure 2.2.

The gossip structure is fixed ex ante- who tells whom about their games is fixed. However, players are randomly assigned opponents so their *objects* of gossip vary. Therefore, even though person 1 always learns about the games of player 2, 1 also receives information about the opponents that 2 is randomly assigned to play.

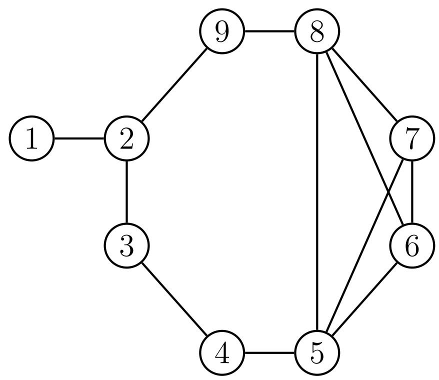

Figure 2.1: Example communication network with 9 players.

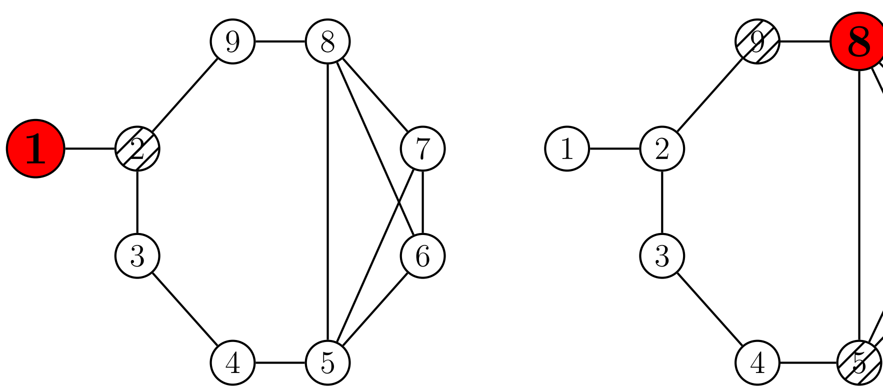

Figure 2.2: Gossip's reach in $t+1$ when it starts with player 1 (left) and when it starts with player 8 (right).

The infinitely repeated game G proceeds as follows: at the beginning of each round, nature creates a roster of random pairings for both groups. Players in each group observe only their own pairing on the roster. When a player is matched with an in-group member, the players know each other's identity. When a player is matched with an out-group member, the players know only that they are playing someone in the other group. Pairs play one round of the prisoner's dilemma. Gossip is spread at rate r, ending the round, and the next round begins.

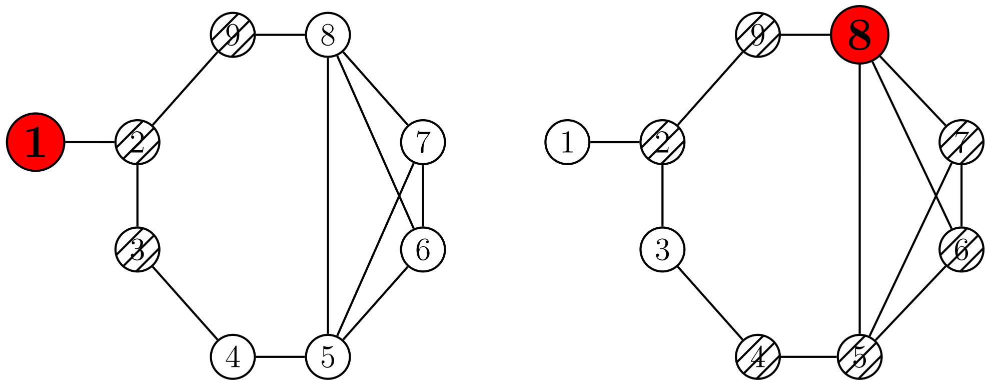

Figure 2.3: Gossip's reach in $t + 2$ when it starts with player 1 (left) and when it starts with player 8 (right).

The Model: Equilibrium Strategies

When the network is incomplete, players may be informed about only a strict subset of the history of play in the game. Cooperation can be sustained if players cope with missing information by following a "presumption of innocence" strategy. When a player has received gossip (or has observed) that another player has defected, that player should follow the same punishment rules as in the original model in Fearon and Laitin (1996). When players have not received information that another player has defected, players are to presume the other is cooperative. Others punish each other for deviating from the prescription, which includes punishing without confirmqation of a defection (i.e., for presuming guilt). Players' strategies are then a mapping from their personal experience; their received, unexpired gossip; the communication network; and their random pairing into an action in $\{C, D\}$. Consider the following generic individual strategy:

Definition 2.1 (**Network Tit-For-Tat** $NWTFT$)**.** *Always punish a player (play* D*) known to be in punishment phase. Always cooperate with a player (play* C*) not* *known to be in punishment phase.*

Punishment phase is a status defined by the strategy profile below- in general, being in punishment phase is undesirable and the threat of entering punishment phase is what drives players to cooperate. Here, a player can "know" because of experience in his own rounds, or because of received gossip. A player is "not known to be in punishment phase" by a player i if i has not experienced or heard gossip that the player is in punishment phase. Now consider a strategy profile in which players play NWTFT according to the following guidelines:

Definition 2.2 (**Network In-Group Policing** σ^{NWIGP}**).** *All players play NWTFT using the following definitions of status: all players begin as cooperators (not in punishment phase). A player enters (or reenters) punishment phase for T^p periods when that player (1) defects against an out-group member, (2) defects against someone not known to be in punishment phase, or (3) cooperates with someone known to be in punishment phase. A player i is known by his opponent to be in punishment phase when his opponent was the victim of (2) or (3) committed by i in at least one of the past T^p rounds or has received a message that i committed (1), (2) or (3) in at least one of the past T^p rounds.*

Note first that punishment here takes the form of capitulation (playing C against a punisher playing D) as it does in Fearon and Laitin (1996) and Calvert (1995). That defectors may be required to atone for their offense by capitulating for a certain number of rounds has the nice property that players like to punish defectors, and squares with real world observed punishment regimes (Harbord, 2006).

Second, note that the above strategy profile is really a *set* of strategy profiles, each with a particular length of punishment phase, T^p. Just as a group is assumed to coordinate on the strategy $NWTFT$, the group is assumed to coordinate on a particular length of punishment phase to use. How long we could expect punishment to last depends on how salient the offense is– whether it becomes old news quickly or sticks in the group's memory– and how difficult the group is to keep in line. Below I show that some groups have members with large incentives to cheat, and such groups may anticipate this problem and coordinate on a long, severe punishment to threaten. Noting the particular difficulties the Navaho face in keeping members cooperative under "minimal government" (Colson, 1974, p.41), Kluckhohn notes: "[a] society like the Navaho *must* have harsh sanctions against physical aggression" (Kluckhohn, 1944, p.53). Below I formally derive this need.

Given the desirableness of equilibria in which cooperation does not break down forever in the face of an accidental defection, we may be interested in the shortest possible punishment phase that can support cooperation. While in the original it is the case that if cooperation can be supported at all it can be supported for a single round of punishment, here that is not always the case. More on this below. The strategy profile inflicts the same punishment on all defectors committing the same offense. Such egalitarian punishment regimes are typically favored in the real world. In Chapter 3, the efficiency of discriminatory punishment in which T^p varies by person is considered. Here, all punishment regimes are egalitarian; like offenses earn like punishment phases.

Third, not only does this strategy profile presume innocence, which, as mentioned above, is both realistic and useful in uncertain environments, the strategy profile also saves players from the taxing calculations of trying to figure out information they have not received. While the problem of inference on a network can be arduous to say the least[14] (Goyal, 2007), here players have no incentive to learn information beyond what they experience and what they are told. If they act on information other than what they are told, either that information confirms innocence, in which case the player fares no better than not knowing, or that information reveals guilt, but punishing based on this inference rather than received gossip earns punishment. The strategy encodes a natural tendency to avoid hard calculation, especially when those calculations could also presume guilt without confirmation.

To follow σ^{NWIGP}, a player must keep a record of deserved rounds of punishment of all other players for T^p rounds, based on his own experience and the gossip he has received. The natural point at which the gossip expires is then T^p rounds after it is first sent. A player's record of deserved punishment may undercount the number of players who *would* be in punishment phase if everyone were perfectly informed of everyone else. Since messages are truthful, though, he will not over-count the number in punishment phase. In this way an incomplete network with slow enough gossip transmission attenuates punishment because some players will not be aware that punishment is deserved at the time they're assigned to play the person who defected. To ensure full cooperation, all prospective defectors must expect a large enough punishment.

[14] Arduous for the researcher, let alone for the player in the game supposedly making these calculations on the spot.

A Closer Look at the Information Structure

All relevant information that i has about in-group members at time t can be captured by the number of rounds that i knows are left in a player's punishment phase, so player i's record of history at time t can be though of as a vector $\mathbf{s_{i,t}}$ of length n with entries $s_j \in 0, \ldots, T^p$ where $s_j = 0$ indicates that player j is not in punishment phase from the perspective of player i. Out-group members never incur punishment by players outside their group, so trivially always deserve 0 rounds of punishment from the other group members. Entries are 0 unless one of i's rounds or received messages provides information to the contrary. Messages transmit information that can be used to determine the number of rounds left in a player's punishment phase. Since messages are truthful, no player mistakenly records a number of punishment rounds left that is too high. If the network is incomplete, though, a player *can* mistakenly record a number of punishment rounds left that is too low (i.e. is lower than it would be if the network were complete). If player i defects in time t, player j will only know about this if player j was the victim, or is a neighbor to i or to the victim, or is a neighbor to a neighbor of i or the victim and gossip spread quickly enough, and so on. The total number of players eligible to punish in $t+1$ depends on the structure of the gossip network.[15] With probability $1-p$, the defector will be assigned an in-group opponent in $t+1$; the more neighbors he or his victim had, the greater the probability he will play someone in $t+1$ who knows he deserves punishment, and hence the greater the probability that he will incur this round of his punishment

[15] As discussed below, when the punishment phase is longer than a single period, or news spreads more rapidly, characteristics of the network other than degree become important too.

phase.[16]

One way to conceptualize the information that players learn and remember about other players is to imagine a set of "deserved punishment counters." Imagine that each player keeps a set of n devices. Each device stores the number of rounds left in his own and a particular other player's punishment phase. Assume the devices are generous: the default displays are 0. That is, in the absence of evidence of defection, players assume others are cooperators and do not deserve punishment. Each player is responsible for setting and resetting his set of counters, which he does when he learns that someone has defected.[17]

Counters display 0 for all players who have not defected. If perfectly informed, counters display T^p for any players who defected last period, $T^p - 2$ for players who defected three periods ago, and so on. Each player updates his set of counters according to information he learns through the gossip network, and in each period the counts automatically decrease by one. Players who experience or hear gossip about defections set their displays accordingly. If player i defects in t, his victim, his neighbors and his victim's neighbors observe/hear about the defection and set their punishment counters for player i. When the network is complete, everyone is a neighbor of the defector (as well as a neighbor of the victim), so everyone (and hence the next opponent) will always have accurately-set counters. Players condition their

[16]It follows that defections against in-group members are weakly more likely to be punished than defections against out-group members since an out-group victim does not gossip with other in-group members. If the network is connected (so that there are no isolates), defections against in-group members are strictly more likely to be punished.

[17]Think of this as a set of world clocks that must be set by the clocks' owner in a world with erratic daylight savings rules that must be learned first-hand or from friends. Or, less imaginatively, a player remembers a vector of counts that correspond to the number of rounds left in a punishment phase.

punishment on the readings of their set of counters, so someone deserving punishment will be punished in every round of T^p when the gossip network is complete.

In an incomplete gossip network, some may not hear about a defection immediately and so will leave their counter at the previous value, possibly 0. Gossip about a defection travels directly to neighbors of the victim and the defector, and spreads to their neighbors, and to their neighbors' neighbors, and so on at some rate r. When the gossip network is complete, the fact that information flows in this way is irrelevant; all that matters is that as a consequence, everyone always knows about all defections. When the gossip network is incomplete, the fact that information flows this way is relevant since different patterns of connection, different shapes of the gossip network, determine who knows enough to punish. The next section formalizes the way punishment, and hence cooperation, depends on the communication network.

Full Cooperation in Equilibrium: The Role of the Network

The strategy profile σ^{NWIGP} instructs players to punish members of their own group who are known to be defectors, and, if severe enough, this punishment threat entices players to be cooperative. Even when the gossip network is incomplete, this strategy profile can produce full cooperation in equilibrium under certain conditions. To keep everyone cooperating, we can look to the incentives of the players that are the most difficult to keep in line. A player's profit from deviating depends on his probability of facing future punishment when paired with an in-group member.[18] Let

[18]By construction players are never punished by out-group members. All out-group pairings are supposed to be cooperative, and when someone deviates, it is the responsibility of the in-group to punish the misbehavior.

$N_{i,rt}$ be the neighborhood around i of size rl, i.e. the set of players reachable from i in rl degrees or fewer. Then for communication rate r and punishment length T^p,

Definition 2.3 (**Punishment Probabilities**). *The probability that i will be punished in $t+l$ by an in-group member for a defection against an out-group member in t is*

$$z_{i,rl}^{out} = \frac{\#N_i^{rl}}{n-1}$$

and the probability that i will be punished in $t+l$ by an in-group member for a defection against in-group member j in t is

$$z_{i,j,rl}^{in} = \frac{\#(N_i^{rl} \cup N_j^{rl})}{n-1}$$

Based on the way information spreads through the communication network, the probability of punishment for out-group offenses depends on the number of players reachable from i in rl degrees or fewer. The probability of punishment for in-group defenses depends on these players as well as those reachable from the in-group victim. Information from out-group victims never reach in-group members since by assumption no communication ties span the two groups.

Clearly as more time passes since the defection (l increases) or information spreads more rapidly (r increases), these probabilities become weakly larger. This observation is taken up below.

Now we can characterize the most likely defections– who are the most tempting targets and who are the most likely offenders.

Definition 2.4 (**Most Tempting Opponent**). *A player's most tempting opponent for $t+l$ is the opponent against whom a defection in t would be most detectable in*

$t+l$ given gossip rate r. All out-group opponents are equally tempting to player i. Player i's most tempting in-group opponents for $t+l$ are the j such that

$$\arg\min_j \{z^{in}_{i,j,rl}\}$$

That is, i's most tempting in-group opponents for $t+l$ are those which minimize the probability that i is punished in $t+l$ for his in-group offense against j in t given gossip rate r.

Each player is most tempted to defect when they play their most tempting opponent. The player facing the largest temptation can be identified:

Definition 2.5 (Least Punishable Player LPP). *The* ***least punishable player against the out-group*** *is the player who faces the lowest probability of punishment in $t+l$ for his defection against the out-group given gossip rate r, i.e. is the i such that*

$$\arg\min_i \{z^{out}_{i,rl}\}$$

and the ***least punishable player against the in-group*** *is the player who faces the lowest probability of punishment in $t+l$ for his defection against his most tempting in-group opponent given gossip rate r,*

$$\arg\min_i \{z^{in}_{i,MTO,rl}\}$$

Call the probability that the least punishable player will be punished in $t+l$ for a defection against his most tempting out-group opponent in t $z^{out}_{min,rl}$ and the probability that the least punishable player will be punished in $t+l$ for a defection against his most tempting in-group opponent.in t $z^{in}_{min,rl}$.

Players who face especially small values of $z^{out}_{i,rl}$ and $z^{in}_{i,MTO,rl}$ can be said to be **peripheral**. Relatively little information reaches or is sent by them.

Now we can turn to the conditions for full cooperation in equilibrium. Because the game is an infinite horizon multi-stage game with imperfect information, subgame perfection has no bite so the more restrictive sequential equilibrium is more appropriate. Strategies must be sequentially rational given they way players form beliefs over their missing information, and beliefs must be consistent (Kreps and Wilson, 1982).

Proposition 2.1 (Full Cooperation). σ^{NWIGP} *is sequentially rational for game G with networks $g_A = g_B$ iff*

$$\delta^{T^p} \geq \max\left\{\frac{\alpha - 1}{(1-p)z^{out}_{min,rT^p}(\beta+1)}, \frac{\beta}{(1-p)z^{in}_{min,rT^p}(\beta+1)}\right\}$$

and

$$p < \min\left\{\frac{z^{in}_{min,rT^p}(1+\beta) - \beta}{z^{in}_{min,rT^p}(1+\beta)}, \frac{z^{out}_{min,rT^p}(1+\beta) - \alpha + 1}{z^{out}_{min,rT^p}(1+\beta)}\right\}$$

where r is the speed of information spread, T^p is the length of punishment, and z^{out}_{min,rT^p} and z^{in}_{min,rT^p} are the probability that the least-punishable defection will be punished by an in-group member at the end of the punishment phase (in $t + T^p$) for defecting in t against an out-group member and in-group member, respectively, as in Definitions 2.3 and 2.5.[19] The Appendix contains the proof of sequential rationality, and discusses consistent beliefs which extend the behavior to be a sequential equilibrium.

[19] As the Appendix makes clear, the final round of the punishment phase is the relevant period because the binding player is one considering defecting a second time in a row, adding an additional round of punishment in $t + T^p$. If he can be made to cooperate, someone considering defecting a first time can be too.

Note that the conditions reduce to the condition in Fearon and Laitin (1996) when the network is complete (because then $z^{out}_{rT^p} = z^{in}_{rT^p} = 1$). The game in Fearon and Laitin (1996) can then be considered a special case of the more general game here.

All relevant features of the network- its size, its topology- are captured in the single parameter z.

Proposition 2.2 (**Network Effect**). *All features of the network impact the prospects for a fully cooperative equilibrium through the parameters*

$$z^{out}_{min,rT^p} = \min_i \left\{ \frac{\#N_i^{rT^p}}{n-1} \right\}$$

and

$$z^{in}_{min,rT^p} = \min_{i,j} \left\{ \frac{\#(N_i^{rT^p} \cup N_j^{rT^p})}{n-1} \right\}$$

where $N_i^{rT^p}$ is the set of players fewer than rT^p degrees away from i, and $\#(N_i^{rT^p})$ is the number of such players. In network terms, $N_i^{rT^p}$ is i's neighborhood of size $N_i^{rT^p}$. So z^{out}_{i,rT^p} is the proportion of in-group members reachable from i in rT^p degrees; z^{out}_{min,rT^p} is the smallest such proportion in the group. Likewise, z^{in}_{i,j,rT^p} is the minimum proportion of in-group members reachable from either i or j in rT^p degrees and z^{in}_{min,rT^p} is the smallest in the group.

For fixed punishment length and gossip speed, z is effectively a measure of how un-knowable the least knowable players are, i.e. how peripheral they are. To take a simple example, when $r = 1$ and $T^p = 1$, Figure 2.4 shows z^{out} for all people in the group.

Likewise, Figure 2.5 shows the minimum punishment probability each would face for defecting against their most tempting in-group opponent. The player with prob-

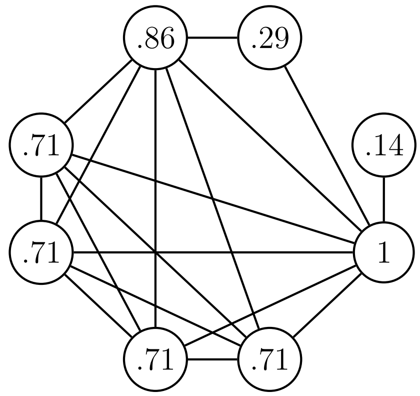

Figure 2.4: z^{out}_{i,rT^p} for each i in the group with $T^p = 1$ and $r = 1$.

ability .14 in Figure 2.4 faces such a low probability relative to the rest of the group because his interactions with the out-group are known by relatively few in-group members. He has only 1 neighbor, and if information spreads slowly enough and punishment expires quickly enough, only this one neighbor ever has a chance to punish him. Other players are less peripheral because their actions are observed by relatively more in-group members.

In Figure 2.5, the same player again faces the lowest punishment probability for his most tempting in-group offense, but is tied with another player who also faces probability .43. This is because in-group defections can be observed by both the victim and the offender. To defect against the in-group, the most peripheral player must select an opponent who is also known by some. Here, he is least punishable when he plays the other relatively peripheral player, the one connected to only two others, one of whom is a shared neighbor. His least punishable defection is then known about and punishable by 3 others.

Proposition 2 is noteworthy for two reasons. First, the network matters only insofar as it affects the probability that the least-punishable defector will be punished

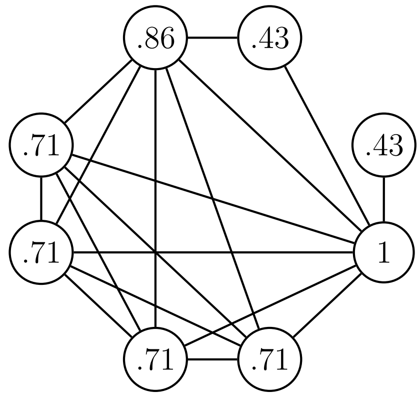

Figure 2.5: $\min_j z^{in}_{i,j,rT^p}$ for each i in the group with $T^p = 1$ and $r = 1$.

at the end of the punishment phase. This is significant because network modelers often encounter a stumbling block when deciding how to meaningfully vary generic networks since networks can vary along many different dimensions (do we vary degree or centrality or diameter or...). The work-around is often to restrict attention to one class of networks, like the class of Erdos-Renyi random networks, so that a single parameter can capture the difference between networks, like the probability that two nodes are linked. Here, we can consider the full universe of networks since the model provides us with a single parameter with which to compare them in terms of full cooperation: the probability that the least punishable player will be punished for defecting at the end of the punishment phase.

Second, z functions as a measure of how likely punishment will keep each player cooperating. Since the approach here admits individual-level heterogeneity on this measure, it formalizes one interpretation of trustworthiness batted around in the debate among sociologists over how to measure and operationalize social capital and trustworthiness (Kramer, 1999; Hardin, 2002; Burt, 2005). There seems to be consensus that the concept has something to do with social networks, but mapping off-the-shelf network statistics onto the concepts has led to confusion and disagreement (Borgatti, Jones and Everett, 1998; Glaeser, Laibson and Sacerdote, 2002; Durlauf, 2002; Cook, Hardin and Levi, 2007; Barr, Ensminger and Johnson, 2010). The measure produced from the model presented here has the advantage that it is theoretically grounded. Its meaning is clear because it is derived from consciously-chosen assumptions. Players are more trustworthy due to reputation mechanisms when they face higher likelihood of punishment for their offense which is governed by their position in the social network.

Improving Cooperation

The comparative statics from Propositions 1 and 2 can be used to assess problems with full cooperation and potential improvements. First, take the length of the punishment phase, T^p, which features on both sides of the equilibrium condition. When the network is complete, if there exists a fully cooperative equilibrium, it will hold for the shortest punishment phase, $T^p = 1$. When the network is incomplete, this is not necessarily true. Each round that ticks by gives more people the chance to have the news reach them and so increases the number of possible punishers. This means that a longer punishment phase increases the probability of punishment at the end of the punishment phase, which makes the equilibrium condition easier to satisfy. However, a longer punishment phase also delays the final round of punishment, making it less significant to impatient players. This makes the equilibrium condition harder to satisfy. We can then make the following generalization:

Proposition 2.3 (Length of Punishment). *For sufficiently large δ, the equilib-*

rium condition for full cooperation is easier to satisfy as T^p increases. In fact, in a connected network[20]*, $\exists d$ such that $\forall \delta \geq d$, $\lim_{T^p \to \infty} z^{out}_{rT^P} = 1$.*

The proof is immediate from the definition of z. That is, if players are sufficiently patient, cooperation is increasing in the length of the punishment phase. This is because the more time that passes, the farther information can spread through the network. A longer punishment phase contains periods in which weakly more people are aware of the defection and can participate in punishment. If the network is connected, the probability of punishment approaches 1 as the punishment phase becomes long enough. This suggests one avenue for compensating for sparse communication networks: if players are patient enough, sparsely connected groups will behave more like densely connected groups if they have a longer punishment phase.[21]

The next proposition is the analog for gossip speed:

Proposition 2.4 (Speed of Gossip). *For sufficiently large δ, the equilibrium condition for full cooperation is easier to satisfy as r increases. In fact, in a connected network, $\exists d$ such that $\forall \delta \geq d$, $\lim_{r \to \infty} z^{out}_{rT^P} = 1$.*

This proof is also immediate from the definition of z. Increasing the speed of gossip spread (by, say, improving communications technology) improves prospects for full cooperation by the same reasoning as increasing the length of punishment. The more quickly information spreads, the more quickly a large number of people

[20] A connected network is one in which there exists a path of finite length between any two players, not to be confused with a *complete* network which contains all possible links. A complete network is connected but a connected network need not be complete.

[21] Of course, relying on the patience of agents and infinite length punishment strategies can be unappealing for pragmatic and robustness reasons. See Gintis (2004) for a criticism of such strategies.

become aware of offenses and can punish. Moreover, if players are patient enough, the punishment probability attains its maximum value when news spans the whole network before the next round of play. This is possible if the network is connected and rT^p is large enough. Let $diam(g)$ be the diameter of network g.[22] Two corollaries follow.

Corollary 2.1 (When Gossip Saturates a Network). *For sufficiently large δ, in a connected network, all gossip will reach everyone before the next round of play if rT^p is bigger than the diameter of the network. That is, $\exists d$ such that $\forall \delta \geq d$, $z^{out}_{rT^P} = 1$ when $rT^p \geq diam(g)$.*

Hence, *connected* networks can behave like complete ones *if* patience, punishment length and communications technology are right. This is not possible for unconnected networks, discussed in a later section.

Corollary 2.2 (Minimum Punishment Length). *There exist sets of parameters that satisfy the conditions for full cooperation only if $T^p \geq \tau$ for some $\tau \in \mathbb{N}$, $\tau > 1$.*

In other words, when the network is incomplete, full cooperation may require players to threaten a punishment phase strictly longer than 1 period.

The significance of this finding should not be overlooked. Assuming a complete network (as in the original model) precludes the possibility of punishment threats longer than a single round, or grudges that linger beyond a single chance to express them since if any punishment threat works, a threat of length one would. Of course

[22] The diameter of a network is the length of the longest shortest path connecting any pair of players, i.e. if we list for every pair of players the smallest number of links in a path from one player to the other, the diameter is the largest number in the list. If two players can't reach each other (i.e. the network isn't connected), the diameter is said to be infinite.

we do observe lasting grudges and harsher punishment threats in the real world (expulsion, banishment, persistent feuds, etc.). This framework gives us a means of understanding why, and of understanding variance in punishment length. Groups with isolated players (small zs) and slow-spreading gossip (small r) should be the ones expected to have longer punishment (larger T^p)– groups in environments favoring isolation, with limited ability to communicate, and so on. Contrast blood feuds in which a single instance of homicide is committed in punishment with punishments for cheating in trade that can last many iterations of 'play' (Boehm, 1984; Greif, 1989). Blood feuds are public in a very densely connected social network; trade is less visible and occurs in a larger, less dense communication network (Harbord, 2006). More generally, we should expect punishment phases of short duration to exist only among groups with very dense gossip networks and/or excellent communications technology.[23]

Comparing Groups' Cooperativeness

Proposition 2.5 (Nested Networks). *For a given population N, for any $g'(N) \supset g(N)$, $z(g'(N))^{in}_{min,T^p} \geq z(g(N))^{in}_{min,T^p}$ and $z(g'(N))^{out}_{min,T^p} \geq z(g(N))^{out}_{min,T^p}$.*

The proof is immediate from the definition of z. That is, if one network is a subnetwork of another, the probability of punishment at the end of the punishment phase is lower for the subnetwork and hence cooperation is more difficult. Another way of stating this result is: given any network, adding a new link will always weakly

[23]There is good reason lasting feuds break out between people in outskirts. This issue is taken up further in Chapter 3. Feuds here are out-of-equilibrium behavior which we might explain by a mistake or tremble; Chapter 3 shows that some feuding can be equilibrium behavior as well.

increase cooperation.

This suggests that, for a given group, attempts to forge new meaningful lines of communication between members can be useful because it increases the reach of news that allows people to punish wrongdoing. While manipulating the communication network poses challenges, changes are not impossible, and examples of this improvement are discussed below.

However, simply because adding a link to an *existing* network weakly improves cooperation, that does not mean that any network with more links will be more cooperative than a network with fewer links. Two networks cannot be compared by simply counting links.

Proposition 2.6 (More Is Not Necessarily Better). *The topology of the network matters. If network $g'(N)$ contains a larger number of links than $g(N)$, this does **not** imply that $z(g'(N))^{in}_{min,T^p} \geq z(g(N))^{in}_{min,T^p}$ or that $z(g'(N))^{out}_{min,T^p} \geq z(g(N))^{out}_{min,T^p}$.*

Having a greater number of links does not guarantee that cooperation is easier. Constructing a counterexample is straightforward– see the two gossip networks in Figure 2.6 below. Each has six members. The gossip network on the left has 11 links, the network on the right has 6. Suppose the punishment phase lasts a single period and gossip travels one link per round. The network on the right is actually better able to achieve full cooperation. To see this, note that the equilibrium condition in Proposition 1 depends on the probability that the least punishable player will be punished. Consider defections against the other ethnic group (not pictured here-for simplicity, we could assume both ethnic groups have identical group structures, though this assumption is not necessary), and suppose the punishment phase lasts a single round and gossip spreads one degree per round ($T^p = 1$, $r = 1$). Player 3 is the least punishable player in the network on the left; if he defects against the other group, only player 4 knows at the end of the round. Since punishment only lasts one round, only player 4 can punish. In the network on the right, all players are equally punishable, so consider player 3 again. If player 3 defects against the other group, both players 1 and 5 can punish in the next round. If players are matched uniformly at random, that means in the denser network on the left, the probability that 3 will be punished, assuming he plays his own group, is $\frac{1}{5}$, whereas the probability that 3 will be punished in the network on the right, assuming he plays his own group, is $\frac{2}{5}$. The less dense network is better able to sustain cooperation here.

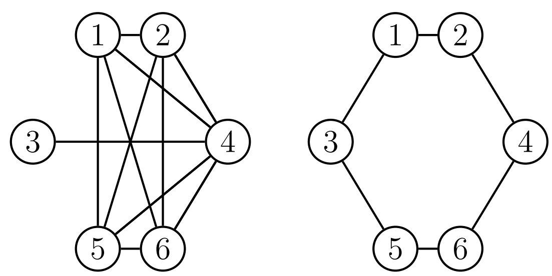

Figure 2.6: Left: 6 players and 11 links, Right: 6 players and 6 links.

Hence, even if a gossip network has a greater total number of links, it may be strictly less able to maintain full cooperation than a gossip network with fewer links for the same number of people. This both affirms that the communication structure should be explicitly considered, and suggests that models brushing aside information transmission issues may be overlooking important variation.

In fact, since the condition for full cooperation depends on the punishment probability of the least punishable player, the relevant statistic to use to compare networks is not the total number of links, but the punishment probability of the least punishable player. The following lemma allows any networks to be compared in terms of full cooperation. Recall from Definition 2.5 that the LPP is the least punishable player and has a probability associated with his least punishable in- and out-group offense.

Corollary 2.3 (Comparing Weakest Links). *All networks with the same probability associated with the* LPP *are equally likely to sustain full cooperation. That is, regardless of how punishable the rest of the members are, if the least-punishable members are equally unpunishable in two different networks, the networks are just as likely (or unlikely) to sustain full cooperation.*

This is simply formalizing the intuition that a group is only as cooperative as its most tempted members (as a chain is only as strong as its weakest link).

This last result also suggests the modifications to a network that would go the farthest toward improving cooperation– add the link that will most dramatically increase the least punishable player's probability– and the breakdowns in a network that would be the most disastrous– those that reduce the lowest probability that any player in the group will be punished.

At this point we know that groups can compensate for an incomplete communication network with longer punishment phases and faster gossip transmission. We know that counting links is insufficient since a network can be more dense and yet be less able to support full cooperation. We also have a means of comparing how easily any two groups can sustain cooperation- we simply look at the probability that their least punishable members will be punished. In the next section I elaborate on these results to examine a few social structures of interest.

Comparing Social Structures: Segregation and Subgroups

Connected networks can be improved, assuming sufficient patience, to the point of resembling a complete network. If instead a group is segregated into subgroups across which there is no regular communication, the prospects for cooperation are dimmer.

When a group is segregated, information gets stuck in subgroups. A player who defects against the out-group or against a member of his subgroup can be know about by at most the number of people in his subgroup.

Lemma 2.1 (Maximum Punishment Under Segregation). *Suppose group A is partitioned into subgroups $a, b, c, \ldots$. The maximum punishment probability that member i of subgroup a could face when defecting against an out-group member or a fellow member of subgroup a is*

$$z_{i \in a}^{outmax} = \frac{\#a}{N-1}$$

and the maximum punishment probability that a member i of subgroup a could face when defecting against an in-group member j from a different subgroup b is

$$z_{i \in a, j \in b}^{inmax} = \frac{\#(a \cup b)}{N-1}$$

Lemma 2.1 follows from Definition 2.3 and is important because it suggests that the fixes above that entail increasing punishment phase (increasing T^p) or improving communications technology (increasing r) or even somehow making players more patient (increasing δ) face limitations in a segregated society. The largest a punishment threat can be pushed is constrained by the size of segregated subgroups and is bounded away from 1. When a defection spans two subgroups, at most people in those two subgroups can know; when a defection is contained in a single subgroup, only people in that subgroup could know.

Consider a stylized example for illustration; Figure 2.7 shows a group segregated into three subgroups, and each is maximally connected (each is a clique).

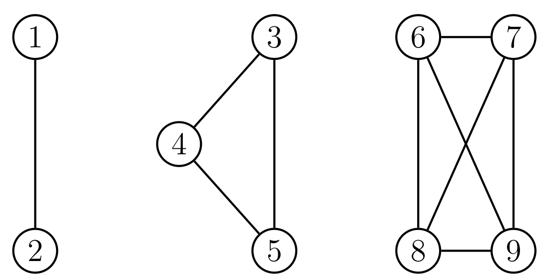

Figure 2.7: Example group segregated into three subgroups.

In Figure 2.7, members of the left most subgroup face at most punishment by one other for defecting against the out-group, for a probability of $\frac{1}{8}$. Members of the same subgroup can also most easily get away with an in-group defection if committed against the other member of their subgroup, also with probability $\frac{1}{8}$. No increase in punishment length or boon to communications technology will increase this probability (while maintaining segregation), let alone drive it to one. If soemone from the left subgroup defects against someone in the middle subgroup, the probability of defection is $\frac{1}{2}$.

Because of this limitation, a few additional results follow. Since the binding constraint to full cooperation is the least punishable player and since the most a player can be punished is limited by the size of his subgroup, the smallest subgroup poses the greatest danger. While it is possible that within a larger subgroup there exists a very isolated player who would be first to defect, we can say that employing fixes to punishment length and communications technology have the least effect in the smallest subgroup. In the best case scenario for full cooperation (i.e. after fixes have been undertaken), the smallest subgroup is the most dangerous. The following corollaries rely on this intuition.

Corollary 2.4 (**The Dangers of Small Strata).** *All else equal, the smallest subgroup of A is the most difficult to make more cooperative.*

Corollary 2.5 (**The Dangers of More Strata).** *All else equal, the greater the number of subgroups in A, the more difficult is full cooperation.*

The above reasoning suggest that adding links to segregated groups can be especially profitable if they span two subgroups. A link that spans two groups increases the maximum punishment of the least punishable, improving prospects for cooperation.

Corollary 2.6 (**The Benefits of Connecting Strata).** *The largest-impact additional link is the one that spans the two smallest sub-groups.*

In the above example, a link connecting someone in the left subgroup to someone in the middle subgroup increases the minimum out-group and in-group punishment probability to $\frac{1}{2}$. No other additional link would be as helpful: connecting the middle to the right subgroup leaves the minimum probabilities at $\frac{1}{8}$, and connecting the left to the right subgroup only boosts the minimum probabilities to $\frac{1}{4}$ (because the unchanged middle group becomes the least punishable).

Comparing Social Structures: Groups with Peripheries

Next, consider a group type in which some people are less embedded in the group–less information reaches them and less is known about them. Such groups may be communities with outskirts or with brand new members who have not had a chance to form connections with many others. The following corollary suggests that such groups have more problems with full cooperation than groups with less remote outskirts (or more quickly-assimilated new members).

Corollary 2.7 (Dangerous Peripheries). *Groups with more peripheral members face the greatest difficulty with full cooperation.*

This follows straightforwardly from Proposition 2.1 and Corollary 2.3. Consider a comparison of two stylized in-groups:

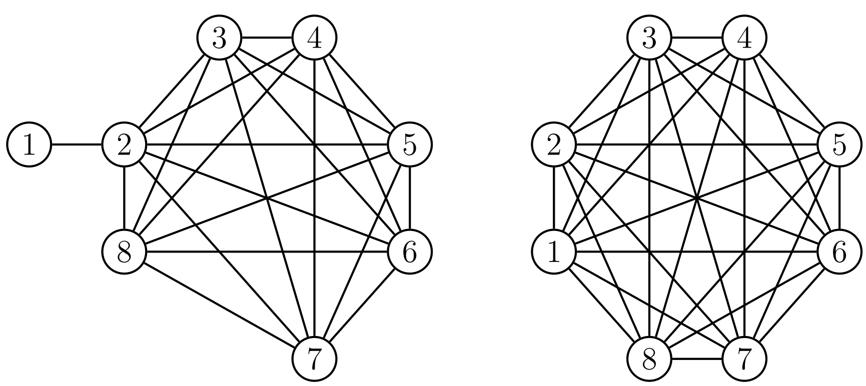

Figure 2.8: Two example groups with 8 members; the group on the left has a peripheral player.

The group on the left has one member, player 1, who is difficult to keep in line, especially with short punishment phases. Groups with peripheries are constrained by their peripheral members. In fact, it can be shown that there exist equilibria in which the peripheral members defect despite incurring punishment while the *rest* of the group is cooperative.[24]

It should be clear that group size *alone* is not a sufficient statistic– the impact of a change in group size is ambiguous without specifying how the structure of connections between people changes with size. There may be a relationship between size and cooperation, but it works through the structure of links that size permits. The story here is that we might expect small groups to be better at fully cooperating, not because of size per se, nor because of the relative density of social ties, but because of the lack of truly peripheral members. Small groups tend to be so integrated that no part of the group is left in the fringe facing low punishment probabilities. Not only does this identify what about the small size matters, but it allows for a more compelling explanation for why small groups often try to stay small. Small groups would do well to reject new members not because the new size of $n + 1$ would be so much worse, but because the new member is likely to be less integrated into the social structure. New members mean peripheral members which can wreak havoc on an otherwise fully cooperative equilibrium, both by acting non-cooperatively or by serving as a tempting target for non-cooperative acts. Fafchamps (2002) notes that in the manufacturing sector in Ghana, Kenya and Zimbabwe, better connected groups tend to be less welcoming to newcomers. The same sort of exclusivity of small, well-connected groups might explain punishment behavior in the inter-caste experiments of Hoff and Fehr (2010) in which members of the high caste were much more likely to punish offenses against members of their same specific caste than were members of the low caste. If we believe that high caste members are more cohesively linked

[24] A detailed look at partially cooperative equilibria and peripheries cheating in equilibrium can be found in Chapter 3.

and enjoy greater opportunities for communicating, then we would not be surprised that this group is more likely to express a proclivity for in-group bias and exclusion. We should expect exclusion when integrating new members cannot be accomplished quickly and thoroughly, and assimilation otherwise.

Comparing Social Structures: Small Worlds and Hierarchies

In addition to peripheries, other aspects of the group structure are telling about the group's cooperative prospects. For instance, it is clear from Corollary 4.2 that the diameter of a network serves as a measure for how 'fixable' it is, and for the demands on communications technology and punishment length. The following lemma summarizes this intuition:

Corollary 2.8 (Low Diameter Groups). *Groups with low diameter are the easiest to improve.*

Lemma 5 follows from Propositions 2.3 and 2.4. The diameter of a network is indicative of the ease with which news from any person can reach any other person; the smaller the diameter, the easier it is to reach anyone. While peripheries add to the diameter of a network, networks with high diameters need not necessarily look like the left network in figure 3.1.[25] Most networks of humans tend to have a lower diameter than networks with something non-human serving as nodes: objects, ideas, etc.[26] Lemma 5 suggests that such networks make sense for humans given

[25] In fact, early study of network models focused on Erdos-Renyi and preferential attachment networks which represented well networks of objects but poorly represented networks of people because the diameters of human networks tend to be smaller than the diameters of these networks.

[26] The class of networks that often describes human networks well is the class of small world networks, with high clustering but low diameter.

evolutionary pressure to sustain cooperation- low diameter networks spread news of defectors more rapidly. In a connected network, a low-diameter network can be made to act like a complete network more easily (and with lighter demands on the length of punishment and communications technology).

The logic about group diameter helps shed light on other classes of group types as well. Consider hierarchical groups. Exactly how the hierarchy is structured can vary, but some features stand out. In one version of hierarchy, the levels are situated so that people in each level communicate with at most people in the single level above and the single level below (in another version, there is no communication across levels). Such an arrangement means that the diameter of the group will always be higher than a different arrangement in which communication can jump across more levels.[27] Hence, the more stratified and rigid the hierarchy, the harder it is to achieve or improve full cooperation. A second feature of hierarchies that is relevant for their cooperative prospects is the extent to which the density of communication varies across levels. It is not difficult to imagine that the lowest level of the hierarchy is also poorly endowed with resources. If this means less time or fewer opportunities or poorer technology for communication, we would expect cooperation to be more difficult within lower levels of hierarchies than higher ones.[28]

Finally, we can say something about the optimal social structures. There is no network more amenable to full cooperation than the complete network, so if links

[27] To see this, note that the smallest possible diameter of a hierarchical group of the form described is the number of levels minus one. The ability to jump more levels decrease the smallest possible diameter.

[28] This comparison is exactly what Hoff and Fehr (2010) find in experimental studies of cooperative behavior within different castes.

are costless, such a network would be optimal. Of course, forming and maintaining connections to people through which meaningful communication can flow in reality entails a cost. Such a cost is not built in here, but assuming a positive cost, we would call the network with the fewest links able to sustain full cooperation the optimal one. From above, we know that small diameters are desirable, as are the absence of peripheral members. If $rT^p = 1$, this implies that the optimal network would be one in which all players had the same degree, a so-called 'regular' network.[29] We might call this kind of network a decentralized one in that no one holds a monopoly on communication- all communicate with the same amount of people.

If instead the group holds longer grudges or communications technology is good enough that news travels more rapidly, then it is no longer necessarily the case that the optimal network (one which doesn't waste links) is a regular one. Now it must be the case that the network is regular in neighborhoods of size rT^p, which, if $rT^p > 1$, does not imply regularity in neighborhoods of size 1.

Corollary 2.9 (Regularity and Optimality). *For groups with better communications technology and/or longer grudges, optimality does not require perfectly equal connectivity. The better the technology/ longer the grudges, the less important is an even distribution of connections.*

[29]This is because the least punishable player must be kept in line. Since $rT^p = 1$, if his degree is high enough, he cooperates, which means all the rest would cooperate if they shared this degree. Any additional links would then be superfluous.

Compensating for an Incomplete Network

The model highlights potential sticking points to a group's ability to cooperate and suggests options to improve cooperation when a communication network is sparse. Given the value of full cooperation, we might expect groups to try to overcome the problems of poor information. Of course, there are natural limitations to a group's ability to recoordinate their punishment institution to increase the length of punishment (T^p), improve the infrastructure and technology that could speed up the spread of information (r), or restructure social relations to forge new channels of meaningful communication between people (alter g). One read is this is a story of expected changes in the face of shocks: if an opportunity to recoordinate arises or new technology falls from the sky, how would group relations change. Or, if a natural disaster wipes out the usual means of communication or external events lead to sour relations between some people and links disintegrate,how is cooperation likely to be a greater struggle.

Despite the difficulty, there are examples of groups enacting changes of their own agency, or designing institutions as expected given their environment. First, take the length of the punishment phase. Lengthening the punishment phase increases both the magnitude of the punishment (because it is inflicted for additional rounds) and the likelihood of eventual punishment (because gossip has more time to travel before it expires). Montenegrin villagers make use of a carefully accounted exchange of homicides as punishment for an initial offense. A single instance of homicide against a single other person is usually called for (Boehm, 1984). Contrast this with the length of punishment that traders may have faced in a merchant guild or trading organization, especially if they faced banishment from the guild and missed the chance to conduct a series of trades (Greif, 1993). Given the tight-knit social structure of Montenegrin villages and public nature of offenses like homicide versus the uncertain, unobservable environment of long-distance trade, the greater need for longer punishment phases among the traders makes sense. The institution in an environment with a presumably dense communication network without peripheral people could employ shorter punishment than the institution in an environment with a sparser network and the presence of very peripheral traders.

Groups have even found creative ways to forge new ties in the communication network in response to uncooperativeness. Ensminger (2001) offers a case in point. The Orma of northwestern Kenya shifted away from family herding and so for the first time needed to hire cattle herders. Since the hired help were often strangers, a serious principal agent problem ensued in which the herders had strong incentives to cheat the cattle's owner through theft or deceit (like selling the cattle but claiming the cattle died to pocket the profit). In response, the Orma literally added social ties by marrying their daughters to the herders. Such additional links were certainly among the most efficient additions since they connected the most peripheral members, the herders, to the center of families. Gluckman (1955) suggests that something similar explains the extent of cooperation among the Nuer, since the addition of social ties gained through travel required for cattle herding in a variable climate increases the observability of offenses and fosters mediation and peace.

Groups with sparse communication networks that are stuck with limited means to manipulate punishment length, communications technology or social ties have also been found to develop institutions that mimic these fixes as closely as possible. Take the villagers in Medieval Iceland who lived without nucleated settlements or dense social structures and little communications technology to speak of (Miller, 1990). Since opportunities to punish an exact offender could be few and news could spread too slowly for a set of people to wield a sufficient punishment threat, villagers resorted to a broader definition of offender. So long as there are negative externalities to being defected against so friends and family bear some cost when each other is wronged, then groups can approach the expected punishment of a complete network by broadening the targets of punishment. In "vendetta transversale," those associated with the offender are legitimate targets for punishment. In Medieval Iceland, "[w]e should not underestimate the extent to which convenience, expedience, or chance affected the choice of victim. One took who one could get. Surely some of the reason Eyvind was such an ideal vengeance target was that he happened to be riding by Hrafnkels farm; the same was true for several other unfortunate travelers in the sagas. Convenience was openly admitted as a basis of selection" (Miller, 1990, p. 203). Such an intsitution approximates making peripheral members more reachable by defining the target to be a large enough set of people that all are sufficiently reachable. When peripheral members can not be drawn in any other way, this kind of institution is sensible.

A Near-Test of the Model

Finally, a novel set of experiments combines trust games in the field with complete social networks data (Barr, Ensminger and Johnson, 2010). A model for future experimental work, the elicited network is a communication network.[30] Players play a trust game and then trusting and trustworthy behavior is related to position in the social network. The authors note the confusion in sociology over which network measures best capture trustworthiness, and settle on two standards of in-degree and betweenness centrality. Based on their description of who turned out to be the most central by these measures– sedentary entrepreneurs, senior elders, heads of large kinship groups, shopkeepers, village headmen, etc.– these players are likely also the least peripheral by the measure presented above in Proposition 2. Their results confirm the in-group variation predicted here, which is that less peripheral members are more likely to be trusting and trustworthy– their interactions are most likely to be cooperative. This set of experiments serves as a plausibility test of the in-group variance predicted here: more peripheral people are harder to keep in line and fellow group members know this.

Conclusion

With this paper I aim to demonstrate the utility of studying groups' networks and to fill a gap in the theoretical literature relating group structure to cooperation. Existing theory has tended to either overlook the question of group structure entirely, or to consider group structure while imposing restrictive assumptions about agents' strategic capabilities, making agents boundedly rational at best or in some cases even nonstrategic automata. Here I present a game theoretic model which overcomes the standard difficulties of incorporating networks by assuming that players

[30] The authors elicited the network by asking "Who do you usually talk to about any kind of problem in this village/workplace" (p. 75).

meet at random, communicate along a fixed network, and follow presumption of innocence strategies. The novel setup makes integrating arbitrary networks tractable and produces analytical solutions useful for comparing networks, individuals within networks, and improvements to groups' cooperative prospects.

The question of how groups of people overcome incentives to cheat (or defect or be violent) and cooperate in the absence of a centralized enforcer is relevant in many contexts. Many societies live outside the reach of a formal, centralized authority.[31] Weak or failed states offer little to no credible enforcement in many contexts.[32] Developing states with weak legal systems are unreliable for enforcing contracts, leaving the economic sector to find other commitment devices.[33] Sometimes states with established legal systems are too backlogged or have too rigid evidentiary rules that the legal system is effectively unavailable for enforcing contracts.[34] And, in some developed countries with developed legal systems, groups find themselves in contexts in which informal enforcement is preferred to the formal option in place, or defections are not technically illegal and so are beyond the scope of formal punishment.[35]

I show that two groups can be compared with their most difficult members to keep in line, the peripheral ones. Comparing groups by counting the total number of links

[31] Consider tribes in the mountainous regions of Montenegro studied by Boehm (1984), or the loosely organized Nuer studied by Evans-Pritchard (1940).

[32] Consider insurgent groups in Uganda studied by Lewis (2010).

[33] See, for example, Dixit (2003).

[34] Dixit (2003) presents the example of India, with a legal system so backlogged that new cases would take decades to even be heard.

[35] Legislatures may be an example of such interactions, in which informal bargains and trades probably do more to keep behavior 'in line' than formal sanctions. Ellickson (1991) presents another example, of cattle ranchers in Shasta County, California, who rely on an informal system to prevent legal infractions rather than the formal legal system.

in their communication networks is *not* a sufficient statistic for cooperation since some groups with more links can face greater difficulty achieving full cooperation. If patient enough, groups can improve prospects for cooperation by increasing the length of the punishment phase, or improving communications technology, options which are not evident when the communication network is assumed to be complete. Changes to the social structure can also improve cooperation, and some groups have been able to make changes like these. Any two networks can be compared simply and I demonstrated some comparisons: hierarchies with more strata and less communication across strata find cooperation more difficult that hierarchies with fewer strata or less rigid communication divisions. Groups with low diameter can be improved most easily, and segregated groups face a stronger barrier to cooperation than unsegregated groups, the more so the more subgroups. In short, the social structure matters because it governs how well decentralized institutions can deliver punishment. Knowing the structure allows us to assess prospects for cooperation and compare groups and individuals within groups.

The results presented here are partial equilibrium results in that they take the gossip network as fixed. An admittedly richer approach would allow the social structure to respond to cooperative or noncooperative behavior to consider a general equilibrium. Taking the social structure as fixed is a first step. More than this, in many contexts, the social structure is effectively fixed or is sufficiently slow moving that this partial equilibrium approach may be more meaningful than a general equilibrium result which assumed the social structure were quickly mutable. Some links in the gossip network may be missing for reasons other than that the two people are not acquainted. (If this were the only reason the link were missing, we could easily imagine that these two people could become acquainted through random interactions so that a link would forge). Some links are missing because the two lack the means to regularly communicate, or physical barriers prevent regular interactions, or deeper cultural or even legal restrictions keep the two from forming a friendship/ gossiping relationship. While friendships can of course form despite these barriers, we might reasonably expect the behavioral strategy to change more quickly or easily than the social structure. Hence, if the social structure or the existing communications technology are sticky, the partial equilibrium approach is the relevant approach for at least some interval of time.[36]

Two trends that are underway make theoretical results about the role of social structure in cooperation especially useful: experiments are becoming more prevalent and methods to directly elicit social networks are improving. This means that direct tests of comparative statics regarding the exact social structure are possible in both the laboratory and the field (and the laboratory *in* the field). In the laboratory, the communication network which governs who learns information about whom can be experimentally manipulated (see Judd, Kearns and Vorobeychik, 2010). Experimental study of cooperation is hardly new, and results increasingly point to players in fact cooperating more in response to punishment threats (Bo, 2005; Dreber, Fudenberg and Rand, 2010). In the field, researchers now face myriad options for eliciting a real

[36]If there is pressure to improve efficiency, we might think of the results pertaining to cheating in equilibrium results explaining at least the short term after a shock to other parameters in the game- payoffs, say, or ease of communication. While in the long run groups may find a way to repair their fully cooperative equilibrium, if these repairs require changing sticky features, the cheating equilibrium will persist for at least a while.

social network, from asking respondents to list friends or pick friends out of a census (see Currarini, Jackson and Pin, 2009), to asking respondents who they communicate with about certain subjects (see Conley and Udry, 2010), to asking how much time respondents spend with other people (see Bellair, 1997), to devising games which reward honest revelation of friends (Rao, Mobius and Rosenblat, 2007), and the list goes on. Field experiments can take players' position in the network as given to directly test results like those presented below (Barr, Ensminger and Johnson, 2010). Not only does this make the study of social structure a testable enterprise, but it also makes a firm theoretical grounding important for interpreting existing results and guiding future research.

Appendix: Sequential Rationality and Sequential Equilibrium (Proof of Proposition 1)

σ^{NWIGP} is sequentially rational at any information set given any specification of beliefs for game G with gossip networks $\mathsf{g}_A = \mathsf{g}_B$ iff

$$\delta^{T^p} \geq \max\left\{\frac{\alpha - 1}{(1-p)z^{out}_{T^p,min}(\beta+1)}, \frac{\beta}{(1-p)z^{in}_{T^p,min}(\beta+1)}\right\}$$

and

$$p < \min\left\{\frac{z^{in}_{T^p,min}(1+\beta) - \beta}{z^{in}_{T^p,min}(1+\beta)}, \frac{z^{out}_{T^p,min}(1+\beta) - \alpha + 1}{z^{out}_{T^p,min}(1+\beta)}\right\}$$

where $z^{out}_{T^p,min}$ and $z^{in}_{T^p,min}$ are the probability that the least-punishable defection will be punished at the end of the punishment phase when the defection is committed against an out-group member and against an in-group member, respectively.[37]

[37] The condition of symmetry between both groups' networks $\mathsf{g}_A = \mathsf{g}_B$ is an unnecessary simplification. When $\mathsf{g}_A \neq \mathsf{g}_B$, both networks must satisfy the above conditions, which means the condition

Parameters $z^{out}_{T^p,min}$ and $z^{in}_{T^p,min}$ depend on the network g and are presented in full below.[38] Each player's strategy in each round conditions only on observable information: whether or not the player knows his partner is a defector.[39] However, for an incomplete gossip network g, at least one player may have incomplete information about the full history of the game. Some players will not punish some defectors since they do not know about the defections. This is unproblematic for the game, since their strategies only instruct them to punish when they know.

A second consequence of observing only a partial history of play is more subtle. To play a best response, players must compare the expected value of playing the candidate equilibrium strategy to the expected value of defecting. These expected values depend on the probability of encountering other defectors, specifically on the probability of encountering other defectors *that the player knows are defectors.* Sequential rationality requires considering this comparison for all histories. It will suffice here to consider this comparison in all states of informedness a player may be in.[40] If a player only knows his state of informedness and not the true history, he can't be sure

binds for the least punishable player in either network. That is, the relevant $z^{out}_{T^p,min}$ is the minimum of both networks' minima, and likewise for $z^{in}_{T^p,min}$.

[38]I drop the explicit dependence on g for easier reading. These parameters could be written $z(\mathsf{g})^{out}_{T^p,min}$ and $z(\mathsf{g})^{in}_{T^p,min}$. The mapping from an arbitrary network g to $z(\mathsf{g})^{out}_{T^p,min}$ and $z(\mathsf{g})^{in}_{T^p,min}$ can be found below.

[39]This information is observable since the network and the rules governing how information spreads are common knowledge. When a player defects, he can determine who will know in each future round. And of course, he knows who he knows about.

[40]This suffices since, in an incomplete network, the history of play may result in all players being defectors, say, but the player faces a limit to the number of defectors he could possibly know about based on his position in the network and the speed of gossip. Another way to say this is that the mapping from histories to informedness is not unique; for some players, different histories will look the same because they will result in knowledge of the same defectors and lack of knowledge of the rest regardless if they really defected or not. Their state of informedness governs their play and others play when paired with them.

what his state of informedness will be next round, or the round after, and so on. We could specify how a player should make a best guess about their future informedness, but it will turn out that the condition that binds does not depend on this guess, so we can leave beliefs about future informativeness unspecified and show players' best responses regardless of how they form beliefs or guesses about the play they do not observe.

I will show that given proposition 1, no player has an incentive to deviate from σ^{NWIGP} for a single stage in any possible history of play and any beliefs about the history of play. Since payoffs are continuous at infinity, preventing single-stage deviations for any history and any beliefs about history ensures no deviations of any length and hence sequential rationality.[41]

The relevant global history up to time t can be fully captured by the punishment status of each player at the beginning of round t. Modifying the approach in Fearon and Laitin (1996), let vector $s_{i,t} = (t_1, t_2, \ldots, t_n)$ characterize the punishment status that i knows all n individuals deserve at the beginning of period t. Based on i's received gossip and experiences, i may not know about every defection which occurred, so i may underrepresent defectors relative to the true global history, and $s_{i,t}$ does not necessarily equal $s_{j,t}$ for player $j \neq i$. $t_{i,2} = 1$ means player i knows player 2 deserves one more round of punishment; $t_{i,2} = T^p$ means that player i knows 2 defected in the last round and deserves punishment for T^p rounds; $t_{i,2} = 0$ means player i does not think player 2 deserves punishment.

[41] This is related to the single-stage deviation principle in Fudenberg and Tirole (1991) typically used to show subgame perfection. Here the only proper subgame may be the entire game, so subgame perfection has no bite.

When i must make a best guess about his future information, as when in t he must predict which (if any) defectors he will know about in $t+1$, his guess about his future information will be denoted $s^*_{i,t+1}$. As discussed, how he makes this guess will turn out to not matter, so we will leave the formation of these beliefs unspecified. The thought experiment used to test seqential rationality assumes that for some arbitrary history, players switch to σ^{NWIGP}. Since players may not observe all other play, they may not know in t which defections they will know about by $t+1$ or by $t+2$ and so on until $t+T^p$. Since strategies are common knowledge and the thought experiment stipulates all switch to σ^{NWIGP} in t, by $t+T^p$ all players should be cooperative. Players only need to make a guess about periods $t+1$ through $t+T^p-1$.

If players are matched with uniform probability to in-group members, the non-zero entries in $s_{i,t}$ determine the probability that player i will be randomly paired with a person i knows to be a cooperator in t if paired in-group. Call this probability $d_{i,t}$, or $d^*_{i,t}$ when based on a guess of future informedness.

A player has a true status, and a status in the eye of his opponent. A player who has defected a number of periods less than the punishment phase ago is a true defector, a player who has not is a true cooperator. A player whose opponent knows he defected a number of periods less than the punishment phase ago is an apparent defector, and a player whose opponent does not know this (possibly because he is a true cooperator) is an apparent cooperator. We must show that no player has an incentive to deviate in any way regardless of what his opponent knows about him in any history for any guess he makes about his own future informedness. Accounting for a possibly incomplete gossip network, there are seven cases of possible deviation to consider.

A player i who has not defected in the past T^p rounds (a so-called "cooperator") can deviate by

(Ic) cooperating with an apparent defector (i.e. failing to punish someone he knows deserves it),

(IIc) defecting against an apparent cooperator,

(IIIc) defecting against an out-group member.

Likewise, a player i who has defected sometime in the past T^p rounds (a so-called "defector") can deviate by

(Id) cooperating with an apparent defector (i.e. failing to punish when he knows punishment is deserved),

(IId) defecting against an apparent cooperator when the opponent knows i is a defector (i.e. when he is an apparent defector),

(IIId) defecting against an apparent cooperator when the opponent does not know i is a defector (i.e. when he is an apparent cooperator),

(IVd) defecting against an out-group member.

For arbitrary set of $s_{i,t}$ $\forall i \in N$, consider incentives to comply with or deviate from the strategy σ^{igpn} in time t given that everyone else plays σ^{igpn} in t .[42]

[42] Recall that leaving the informedness of players arbitrary means we are leaving the history arbitrary. Our task of ensuring that no player has an incentive to deviate in any history is simplified if we think in terms of ensuring that no player has any incentive to deviate for any set of individuals's informedness about play.

Ic

Clearly no cooperator has an incentive to cooperate with an apparent defector. Deviating from σ^{NWIGP} and cooperating with a defector means the cooperator foregoes his punishment bonus $\alpha - 1$, and is also subject to punishment for deviating from σ^{NWIGP}. Hence, proposition 1 trivially implies no cooperator prefers to defect via Ic regardless of what he knows about other players' statuses.

IIc

Consider a cooperator's choice in t when he has received no information in the last T^p rounds that his opponent is a defector. σ^{NWIGP} dictates that both players cooperate, even if his opponent actually is a defector. The cooperator i playing opponent j whose experience and gossip revealed that the statuses are $s_{i,t}$ in t, who believes he will know statuses $s^*_{i,t+l}$ for $1 \leq l \leq T^p - 1$, and who knows statuses will be $s_{i,t+T^p}$ in $t + T^p$, believes complying yields the expected payoff

$$\begin{aligned} 1 \quad &+ \sum_{l=1}^{T^p-1} \delta^l \Big[(1-p)\Big((d^*_{i,t+l}) + (1-(d^*_{i,t+l}))\alpha\Big) + p\Big] + \\ &\delta^{T^P} \Big[(1-p)\Big((d_{i,t+T^p}) + (1-(d_{i,t+T^p}))\alpha\Big) + p\Big] + \sum_{l=T^p+1}^{\infty} \delta^l \end{aligned}$$

where every starred term is a consequence of his beliefs.

Deviating offers i expected payoff

$$\begin{aligned}\alpha \quad &+ \sum_{l=1}^{T^p-1} \delta^l \Big[(1-p)\Big((d^*_{i,t+l}) z^{*in}_{i,j,t+l}|_C(-\beta) + (d^*_{i,t+l})(1 - z^{*in}_{i,j,t+l}|_C) + \\ &(1-d^*_{i,t+l})(1-z^{*in}_{i,j,t+l}|_D)\Big) + p\Big] + \\ &\delta^{T^p}\Big[(1-p)\Big((d_{i,t+l}) z^{in}_{i,j,t+T^p}|_C(-\beta) + (d_{i,t+T^p})(1 - z^{in}_{i,j,t+T^p}|_C) + \\ &(1-d_{i,t+T^p})(1-z^{in}_{i,j,t+T^p}|_D)\Big) + p\Big] + \sum_{l=T^p+1}^{\infty} \delta^l\end{aligned}$$

where $z^{in}_{i,j,t+l}|_C$ is the conditional probability that i's defection against an in-group member j in t will be known by i's opponent in $t+l$ given that i plays someone he knows to be a cooperator in $t+l$, and $z^{in}_{i,j,t+l}|_D$ is the same given that i plays someone he knows to be a defector in $t+l$. A cooperator then prefers to comply with σ^{NWIGP} if and only if, for all opponents j,

$$\begin{aligned}\alpha - 1 \quad &\leq \sum_{l=1}^{T^p-1} \delta^l (1-p)\Big[d^*_{i,t+l} z^{*in}_{i,j,t+l}|_C(1+\beta) + (1-d^*_{i,t+l}) z^{*in}_{i,j,t+l}|_D \alpha\Big] + \\ &\delta^{T^p}(1-p)\Big[d_{i,j,t+T^p} z^{in}_{i,j,t+T^p}|_C(1+\beta) + (1-d_{i,t+T^p}) z^{in}_{i,j,t+T^p}|_D \alpha\Big]\end{aligned}$$

and since in $t+T^p$ all players are cooperators if they play σ^{NWIGP} starting in t, the condition becomes

$$\alpha - 1 \leq \sum_{l=1}^{T^p-1} \delta^l (1-p)\Big[d^*_{i,t+l} z^{*in}_{i,j,t+l}|_C(1+\beta) + (1-d^*_{i,t+l}) z^{*in}_{i,j,t+l}|_D \alpha\Big] + \delta^{T^p}(1-p) z^{in}_{i,j,t+T^p}(1+\beta) \quad (2.1)$$

where $z^{in}_{i,j,t+T^p}$ is $z^{in}_{i,j,t+T^p}|_C$ when $C = N$, everyone is a cooperator.

IIIc

A cooperator deciding whether or not to defect against an out-group member faces nearly the same expected payoffs as a cooperator deciding whether or not to defect against an in-group cooperator, with one modification. A defection against an out-group player is weakly less likely to be known by a future in-group player than a defection against an in-group player. Recall that neighbors of both players in a round observe the game. When both players are in-group members, both players may have neighbors who can watch. When one player is an out-group member, that player has no in-group neighbors to watch. Hence, weakly fewer in-group players learn about a defection committed about an out-group member. This means the probability that i's defection in t will be known in period $t+l$, z_{t+l}, will be weakly smaller when i's defection is committed against an out-group player: $z^{out}_{t+l} \leq z^{in}_{t+l}$. Using the same reasoning as above, the conditions preventing defections of the type IIIc are then

$$\alpha - 1 \leq \sum_{l=1}^{T^p-1} \delta^l (1-p)\left[d^*_{i,t+l} z^{*out}_{i,t+l}|_C (1+\beta) + (1-d^*_{i,t+l}) z^{*out}_{i,t+l}|_D \alpha\right] + \delta^{T^p}(1-p) z^{out}_{i,t+T^p}(1+\beta) \quad (2.2)$$

where $z^{out}_{i,t+l}|_C$ is the probability that i will be punished in $t+l$ for defecting against an outgroup member in t given that he plays an apparent cooperator in $t+l$.

Id

When everyone plays according to σ^{NWIGP}, cooperation with an aparent defector incurs punishment. Consider the decision of defector i paired with j whom i knows is a defector. If j also knows i is a defector, σ^{NWIGP} instructs both to play D. i prefers to not deviate, more so since deviation incurs punishment. If j does not also know that i is a defector, σ^{NWIGP} instructs i to play D and j to play C. i prefers not to deviate, more so since deviation incurs punishment. Hence, proposition 1 trivially implies no defector prefers to defect via Id.

IId

When defector i has not received any information that j is a defector and j knows i is a defector, σ^{NWIGP} instructs i to play C and j to play D. Player i could deviate to avoid losing β, though he will face future punishment. The length of punishment he will endure for this violation will be, as usual, T^p rounds. However, i was already in a punishment phase since he was a defector. His punishment phase resets when he commits a new defection. Suppose he had t_i rounds left in has last punishment phase. The additional punishment he will incur from this new defection will last for $T^p - t_i - 1$ rounds. Player i prefers to comply rather than defect via scenario IId iff, for all j,

$$\begin{aligned}\beta \;\; &\leq \sum_{l=t_i}^{T^p-1} \delta^l (1-p)\left[d^*_{i,t+l} z^{*in}_{i,j,t+l}|_C (1+\beta) + (1-d^*_{i,t+l}) z^{*in}_{i,j,t+l}|_D \alpha\right] + \\ &\delta^{T^p}(1-p) z^{in}_{i,j,t+T^p}(1+\beta)\end{aligned}$$

Deviating is most tempting to player i when $t_i = T^p$ regardless of what he believes about the status of other players or the history of the game. That is, deviating is most tempting to players at the beginning of a punishment phase, so that another defection would only incur one more round of punishment. The difference between complying and deviating hinges on the expected value of that possible additional round of punishment. Hence, player i prefers to comply rather than defect via IId iff, for all possible in-group opponents j,

$$\beta \leq \delta^{T^p}(1-p) z^{in}_{i,j,t+T^p}(1+\beta). \tag{2.3}$$

IIId

When defector i has not received any information that j is a defector and j has not received any information that i is a defector, σ^{NWIGP} instructs i to play C and j to play C. Player i could deviate to earn the larger payoff α instead of 1, but would face future punishment. As in case IId, the additional punishment that this new defection would bring would last $T^p - t_i - 1$ rounds. All players prefer to comply with σ^{NWIGP} rather than defect via scenario IIId iff, for all j,

$$\alpha - 1 \leq \delta^{T^p}(1-p)z^{in}_{i,j,t+T^p}(1+\beta). \tag{2.4}$$

IVd

The decision facing a defector considering defecting against an out-group member is very similar to the decision facing a defector i considering defecting against an in-group member who doesn't know i is a defector. The only difference is that, as in case (IIIc), a defection against an out-group member is observed by weakly fewer in-group members, and hence is weakly less likely to be punished in future rounds. No defector has an incentive to defect via scenario (VId) iff, for all j,

$$\alpha - 1 \leq \delta^{T^p}(1-p)z^{out}_{i,j,t+T^p}(1+\beta). \tag{2.5}$$

Condensing equilibrium conditions

We are left with five inequalities that must hold for all players to preclude any defection. Before finding the binding player, note that since $\alpha, \beta > 1$, $p \geq 0$ and $z \geq 0$, inequality (5) implies inequality (2) and inequality (4) implies inequality (1).

Since $z_{i,j}^{in} \geq z_i^{out}$ for any i and any opponent j, (5) also implies (4) and therefore (1). Thus, player i has no incentive to deviate iff, for any j,

$$\beta \leq \delta^{T^p}(1-p)z_{i,j,t+T^p}^{in}(1+\beta)$$

$$\alpha - 1 \leq \delta^{T^p}(1-p)z_{i,j,t+T^p}^{out}(1+\beta)$$

Hence, no player has an incentive to defect if the above conditions hold for all i and for any opponent j. The conditions will bind for the least punishable player, so to strip some notation, the conditions for full cooperation become:

$$\alpha - 1 \leq \delta^{T^p}(1-p)z_{min,T^p}^{out}(1+\beta)$$

and

$$\beta \leq \delta^{T^p}(1-p)z_{min,T^p}^{in}(1+\beta)$$

where $z_{min,T^p}^{out} = \min_i\{z_{i,t+T^p}^{out}\}$ and $z_{min,T^p}^{in} = \min_{i,j}\{z_{i,j,t+T^p}^{in}\}$.

Rearranging, the conditions become those in Proposition 1. The conditions on p ensure that the probability of cross-group pairings is well-defined.[43] So long as the structure of the gossip network, the parameters of the game, and the players' discount factors meet the conditions in proposition 1, no player has an incentive to deviate

[43] Note that since the binding case is a player considering defecting a second time (and so extending his punishment phase by one round at the end of the phase), a player's beliefs about the interim rounds $t+1$ through $t+T^p-1$ can be left unspecified.

from σ^{NWIGP} in any way given any history of play and any method of guessing what will be known in future periods. σ^{NWIGP} is sequentially rational in every information set for any specification of beliefs for game G with gossip network g. To show that there exists a sequential equilibrium of the game using strategy profile σ^{NWIGP}, all that remains to be shown is that there exists a set of consistent beliefs.

The behavior in equilibrium is of greater interest than the beliefs in equilibrium. The above shows that σ^{NWIGP} forms half of a sequential equilibrium since the strategy is sequentially rational in any information set with any beliefs over nodes in the information set. Consider a set of beliefs, μ, which places probability one on all events expected in equilibrium, probability zero on out-of-equilbrium events, and which are updated according to Bayes' rule. Now consider a perturbation of these beliefs in which players are assumed to tremble with small independent probability $\epsilon > 0$. Given the perturbation, all information sets are reached with positive probability and Bayes' rule pins down beliefs everywhere. Let μ^* be the limiting beliefs derived from Bayes' rule as $\epsilon \to 0$. Since sequential rationality holds for any specification of beliefs given the above conditions, it holds for μ^*. Hence, the assessment (σ^{NWIGP}, μ^*) is a sequential equilibrium.

Appendix: Punishment Probabilities

The probability that player i will be punished l rounds after he defects is a function of the network g, the speed of gossip $r = \frac{Degrees\ Spread}{Rounds\ Played}$, and the matching process.[44]

[44] Whether he is punished by a known cooperator or a known defector is of course also a function of $s^*_{i,t+l}$. In the condition for full cooperation, the only relevant probability is the probability of facing punishment at the end of the punishment phase, i.e. in $t + T^p$, given that everyone else is a cooperator.

Let C_i be the set of players that i believes to be cooperators, and D_i the set of players that i believes to be defectors. Let $N_i^{rl}|_{C_{i,l}}$ be the set of all players i thinks are cooperators in period $t+l$ that are reachable from i in walks of length rl or shorter. Likewise, let $N_i^{rl}|_{D_{i,l}}$ be the set of all players i thinks are defectors in period $t+l$ that are reachable from i in walks of length rl or shorter.

Now the probability that i is punished l rounds after a defection in t against an out-group player given he is assigned to play an assumed cooperator in l, can be written:

$$z_{i,t+l}^{out}|_C = \frac{\#(N_i^{rl}|_{C_i})}{\#C_i}$$

which is the number of assumed cooperators reachable from i in paths of length rl out of the total number of assumed cooperators. The probability when assigned to play an assumed defector is defined analogously.

When defecting against an in-group member, gossip spreads from both the defector and the victim, so the relevant set of reachable players must account for this. The probability that i will be punished in $t+l$ for defecting against in-group member j in t when i plays a cooperator in $t+l$ is:

$$z_{i,j,t+l}^{in} = \frac{\#(N_i^{rl}|_{C_i} \cup N_j^{rl}|_{C_i})}{\#C_i}$$

which is the number of assumed cooperators reachable from either i or j in paths of length rl out of the total number of assumed cooperators. The probability when assigned to play an assumed defector is defined analogously.

In the condition for full cooperation in the above Appendix, the relevant probabilities are:

$$z^{out}_{min,T^p} = \min_i \{ \frac{\# N_i^{rT^p}}{n-1} \}$$

and

$$z^{in}_{min,T^p} = \min_{i,j} \{ \frac{\#\left(N_i^{rT^p} \cup N_j^{rT^p}\right)}{n-1} \}.$$

Chapter 3

Cheating Because They Can: Networks and Norm Violators

Groups can maintain cooperation well without formal governance so long as they are cohesive enough to observe and punish transgressions. Decentralized punishment schemes have been found keeping cattle grazing in the right places, fostering trust among traders in underdeveloped markets, and preventing neighboring coethnics from fighting, to name a few (see Ellickson (1991); Landa (1981); Greif (1993); Fearon and Laitin (1996)). However, sometimes a cattle rancher tresspasses, a trader cheats another trader, or a member of an ethnic group raids the neighboring ethnic group. I present a game theoretic model of decentralized interactions which explains when to expect a partial failure of informal governance, the form the failure is likely to take, and possible fixes.

In particular, I generalize Fearon and Laitin (1996) to account for the way information spreads from person to person within a community. Under such a setup (as in real life), some people have less access to information about other people, and some people's actions are less known by other people. Given this heterogeneity, some people are more likely to prefer to violate cooperative norms of a group– even if they face the same cost to punishment– because they are more likely to get away with their defection. Exactly which people face an incentive large enough to risk punishment and cheat in equilibrium can be determined, as can their most profitable targets. The identity of the cheaters and kind of cheating depends on the structure of communication (the communication network) and the way information is passed from person to person.

This setup allows the comparison of a range of uncooperative behavior under the same framework. For example, whether inter-group conflict or fighting in the outskirts or pillaging by a king or raiding by an outlaw or long-lasting feuding obtains for groups can be analyzed through the lens of groups' communication networks and the way information spreads through them. While surely information transmission is not the only source of variance across communities, it is a mechanism central to theories of informal governance and likely to play a crucial (even if not the only) role in the success or breakdown of cooperation.

I show that members of groups that are relatively isolated– who share few communication channels with other group members and are far from the locus of information– are the problem members and if they are isolated enough they may prefer to defect against the other group or even defect against their fellow in-group members. Contrary to the standard "neighborhood watch" intuition, when the problem players are highly clustered they are more likely to be an even greater problem, defecting against even more people. A group can either threaten an inefficiently large common punishment or an efficient but discriminatory punishment to increase the chance that the problem players act cooperatively. The nature of information spread determines the precise form of in-group cheating– whether the more isolated players are the offenders or the targets, and whether geographic distance matters.

Persistent Cheating

A breakdown of full cooperation can take many forms. I will consider two broad classes, inter-group and intra-group cheating. The act of cheating can also vary, and may be homicide, theft, deceit, norm-breaking, or other uncooperative behavior, depending on the context. The theory presented here serves as a framework under which otherwise disparate phenomena can be understood and compared. I zero in on a mechanism common to a range of types of cheating– the structure of communication– to understand when certain kinds of cheating are possible.

One class of persistent cheating is out-group cheating. In this form, *most* people cooperate with everyone, but the threat of decentralized punishment is insufficient to keep some people from defecting against out-group members. Examples of inter-group cheating range from inter-gang violence to tribe-spanning feuds to colonial v. native conflict (see Venkatesh (2008), Gluckman (1955), McGrath (1987)).

Given the importance of information to punishment institutions, the lack of communication across groups can pose special difficulties for cooperation. I follow Fearon and Laitin (1996) and take up the card case for cooperation, in which there are assumed to be no communication channels that span the two groups.[1]

One setting in which this separation is especially likely to hold is one in which at least one of the two groups was mobile and recently started inhabiting the space. When languages also differ, the separation of the two groups is even more likely to persist. Here I consider Aurora, a frontier town which grew to a population in the thousands from a handful of miners in less than a yearMcGrath (1987). The settler population was predominantly of european descent migrating from the east; the native population was comprised of the Native American Paiute.

Both groups wanted use of the nearby land. Anticipating conflict, they drew up an informal treaty to try to maintain cooperative relations and fair use of the area's resources. Without a strong presence of a governing authority, both groups were left to rely on in-group policing, encouraging their fellow group members to comply. Most did comply with the treaty, but some committed trespassing and homicide against the other group in violation of the agreement. The historical record is rich enough in this case to identify the individual cheaters. I argue below that who cheated and who did not squares with the predictions of the model.

Because information does not spread between the two groups, inter-group cheating is the most likely cheating to occur if groups cannot maintain full cooperation. It is possible, however, that full cooperation is so difficult that groups experience not just inter-group cheating, but *intra*-group cheating as well.[2] Picking back up the

[1]This may not be such a far-flung assumption; analysis of cell phone calls in Belgium reveals a striking 98% of phone calls were made within regional groups and only 2% spanned the groups, suggesting little inter-region communication, at least by cell phone (Blondel, Krings and Thomas, 2010).

[2]This also means that if a group has no interactions with an out-group or if our interest is in a single group, that group can experience internal cheating.

example of the frontier town, I discuss the interactions within the group of miners and show that here, too, *most* interactions were peaceful (homicide and theft were rare), but some were not and again the identity of the cheaters and of their targets can be explained in terms of how information spread among the miners as per the model below.

Intra-group cheating can take many forms, and I show who are the most likely cheaters and victims in various settings. It turns out that the more isolated group members are also the most likely to be the cheaters, *so long as information about the offense is spread by both the victim and the offender.*[3] When information spreads differently, it can be the case that the isolated group members are not the offenders but instead serve as the most tempting target of cheating so that the rest of the group gangs up on them. When the isolated group members are connected to other isolated players so that there is "clustering" among these players, they are more likely to cheat both the out-group *and* their fellow in-group members.

Having a framework which accounts for heterogeneity in how well-integrated in-group members are allows a range of additional questions to be explored. For example, sociologists have long sought to understand when and why feuding between occurs between two groups. Gluckman (1955) identified a puzzling regularity while studying the Nuer: when protracted feuds occur, they tend to be between people who live far apart. The approach presented here is ripe for shedding light on his observation. I take up the question of why cheating may be more likely between people far apart within the framework presented here. I show that even if distance is correlated with social

[3]Information can be spread from the offender if his contacts observe the offense; this intuition doesn't necessarily rely on his admitting and broadcasting his own guilt. More on this below.

distance (so that there are more links in the path that connect people geographically far apart), this is insufficient to produce more feuds between people far apart. It must also be the case that interactions are mostly confined to small neighborhoods and that interactions with people far away are sometimes unobservable to a local neighborhood. If players saw each other at a central marketplace or always encountered each other on one of the player's own turf, the result would not hold.

Sticky Cheating

This is a story of imperfect "relation-based" governance, sometimes called "community enforcement" (see, for example, Dixit (2003), Kandori and Matsushima (1998)). While the heuristic "everyone knows everyone" is often tossed around, real social structures rarely feature meaningful channels of communication between every pair of people in a group. Even Facebook, with very low-cost links, are incomplete (Mislove, 2009).[4] This is all the more true in societies that are large, are sparse, have poor communications technology, or are experience population changes, to name a few. Sometimes a communication network is so incomplete or information travels so slowly that full cooperation isn't possible using decentralized punishment.

The logic of full cooperation in information-rich environments is well-understood (see, for example, Aumann and Shapley (1976), Fudenberg and Maskin (1986), Rubinstein (1979)). The study of cooperation in information-poor environments is catching up, but tends to assume a particular class of communication structure (limiting the types of networks under consideration; see Dixit (2003)), or restrict interactions (see

[4]Patterns of activity between friends on Facebook are even sparser (Wilson et al., 2009).

Ben-Porath and Kahneman (1996)). Here, I depart from the usual approach in imperfect monitoring and focus on a class of equilibria little-studied, asymmetric equilibria in which some cooperation occurs.[5]

When full cooperation is impossible, there may be a next-best equilibrium. I show that equilibria in which partial cooperation persists are possible. The character of these possible equilibria –who defects and against whom– depends on the social structure and on the way information travels.

Chapter 2 identifies possible remedies to restore full cooperation for groups that fail to keep all members in-line. Solutions include increasing the amount of time a deviating group member should be punished, improving the speed and availability of communications technology, and forging new social ties strong enough to foster communication.

Evolutionary models suggest pressure in the direction of cooperation, and so we should expect that, given enough time, all groups who could attain full cooperation would. The two conditions are not to be glossed over- time and ability. None of the solutions is trivial to implement. We might expect some groups to face natural barriers to the ability to restore cooperation and the speed with which this can be attained. Parameters can be sticky, making improvements slow, and less-than-full cooperation can persist in the meantime.

Imagine that a fully cooperative group is exposed to a shock. The shock could

[5] Symmetric equilibria are the popular equilibria to study, often for simplicity rather than more principled reasons. Admitting arbitrary networks means admitting heterogeneity in the amount of communication reaching and leaving players. This heterogeneity opens a class of equilibria especially useful when full cooperation is not possible. That there exists a next-best, almost-fully-cooperative equilibria helps explain the persistence of cheating in the real world.

be something like a natural disaster which disrupts communication infrastructure, development which changes returns to cooperating or defecting, or a demographic shift. In such instances, if making repairs or coordinating new equilibria can't happen quickly enough, groups can become stuck in conditions deeming partial cooperation the best they can do.

Many have observed that a shock to a population can disrupt cooperation. Freudenburg (1986) compares Colorado towns which have a relatively stable population to one which experiences a dramatic population boom. The rapidly growing town featured both a sparser density of acquaintances and more crime. Similarly, Dary (1998) notes the particular danger of frontier towns that were experiencing a population boom, or "boom towns." Mining towns on the frontier often contained little in the way of formal law enforcement, and rapid population change could strain the extralegal means of keeping residents in line (Halaas, 1981, p.85).

The argument presented here is that a rapidly changing population poses dangers because new additions to a community are often socially isolated, at least initially. Newcomers have close social ties with few others and so little information reaches them and little is known about them. The ability to monitor, learn about actions, and carry out punishments is more difficult when some are socially isolated. The problem is not merely that the new people could get away with cheating; in some cases the new people serve as tempting targets of cheating. This squares with a puzzling finding in Freudenburg (1986) that the newcomers are not the only perpetrators of crime even though crime is more prevalent when they join the population.

Under plausible conditions, partial cooperation is possible in equilibrium even when full cooperation is not. If groups experience natural barriers to changing sticky parameters, such equilibria can persist.

Model Setup

The game is a generalization of Fearon and Laitin (1996), modified so that players communicate according to an arbitrary communication network, not necessarily the complete network (implicitly assumed in Fearon and Laitin (1996)).[6] Consider two groups, A and B, each with a set of players $N = \{1, \ldots, n\}$. In each time period, all players play one round of prisoner's dilemma with a randomly assigned opponent. With probability p a player is paired with a member selected uniformly at random from the other group; with probability $1-p$ a player is paired with a member selected uniformly at random from his own group.[7] Each pair plays a single round of the prisoner's dilemma with common payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,1 & -\beta,\alpha \\ \alpha,-\beta & 0,0 \end{pmatrix} \end{array}$$

where $\alpha, \beta > 1$ and $\frac{\alpha-\beta}{2} < 1$. Rounds occur indefinitely. A player's total payoff is then a stream of discounted single round payoffs. Players have common discount factor $\delta < 1$.

[6]See Chapter 2 for a more detailed discussion of the model setup and information transmission process.

[7]For simplicity, groups are assumed to be the same size, so this matching is possible. If groups were different sizes, there would be a pair of probabilities such that when some proportion of group A is matched with group B, this implies some different proportion of B is matched with A. The group sizes will place bounds on possible out-group matching probabilities.

Let $(N, g)_A$ be a fixed simple, undirected communication network for group A with nodes N and $n \times n$ adjacency matrix g such that entry $g_{ij} = 1$ indicates the presence of a link between players i and j, $g_{ij} = 0$ indicates the absence of a link between players i and j. Let $(N, g)_B$ be the same for group B. No links span $(N, g)_A$ and $(N, g)_B$.

Players can perfectly identify and recognize in-group members, but can only identify the group of out-group members. The network $(N, g)_A$ is common knowledge among players in A and $(N, g)_B$ is common knowledge among players in B; players know the shape of the other group's network but not who sits where (i.e. players only know the permutation class of the network of the other group).[8] The roster of random assignments and the actions played in each round are not observable to all players. Instead, this information is revealed via messages to a subset of other players determined by the communication network.

In each period t, after playing one round of prisoner's dilemma, a stage of communication occurs. Information about each game played between two players in each round is packaged in a message m containing the identity of the players (as specific as possible), the time period of the round, the actions of both players, and the motives of the players (relevant motives are determined by the strategy. For the strategy profile σ^{NWIGP} below, the relevant motive will be whether D was played out of punishment or defection.)

Messages about games between same-group members i and j in t perfectly identify

[8] In fact, players do not need to know anything about the other group's network to play their equilibrium strategy. This assumption, that they know the shape of the other group's network, simply makes the arrival at the equilibrium presented here more plausible. This way all players of group A could assess how well group B can police their own and vice-versa.

both players, $m_{i,j,t}$, and are sent to the neighbors of i and the neighbors of j in the communication network, e.g. to $N_g(i)$ and $N_g(j)$. Messages about games between players from different groups cannot perfectly identify the out-group opponent, and so if $i \in A$ and $j \in B$, a message $m_{i,B,t}$ is sent to neighbors of i and a message $m_{A,j,t}$ is sent to neighbors of j. A message $m_{i,j,t}$ expires after T^p rounds, in $t + T^p$. Let r govern the speed of communication so that player i does the following r times before $t + 1$ begins: send an unsent message about i's own game played in t and forward all unsent, unexpired messages to all of i's neighbors. This means that unexpired messages originating at i in time t are received by all players reachable in r degrees or fewer before $t + 1$. Messages are sent deterministically and are not manipulated strategically.[9]

Strategies

First, consider a strategy which is an analog to in-group policing which accounts for a possibly-incomplete network.

Definition 3.1 (**Network In-Group Policing** $NWTFT$)**.** *A player always punishes a player (play D) he knows to be in punishment phase, and always cooperates with a player (play C) he does not know to be in punishment phases, using the following definitions: a player begins as a cooperator (not in punishment phase). A player enters (or reenters) punishment phase for T^p periods when that player (1) defects against an out-group member, (2) defects against someone not known to be in punishment phase, or (3) cooperates with someone known to be in punishment phase. A*

[9] Strategic manipulation of messages is considered in Chapter 4.

player i is known by his opponent to be in punishment phase when his opponent was the victim of (2) or (3) committed by i in at least one of the past T^p rounds or when his opponent has received a message that i committed (1), (2) or (3) in at least one of the past T^p rounds.

In other words, players punish based on what they know. Punishment *would* be T^p rounds of capitulation *if* all opponents in the T^P rounds after the defection knew about the defection. If the communication network is sparse or the reach of communication (r) is small, a player can expect fewer rounds of punishment in some cases, and based on the shape of the communication network, different players can expect different amounts of punishment.

The symmetric strategy profile in which all players play $NWTFT$ results in full cooperation in equilibrium when punishment lasts long enough, news travels fast enough, players are patient enough, and the network is connected enough.[10]

Because there is heterogeneity in the degree to which players actions are known by other players, some face a greater temptation to defect even when they would face the same duration of punishment as all others in the group. Those most tempted to defect might be tempted to play a strategy that earns them no punishment when they play fellow in-group members, but that would draw some punishment by defecting against out-group members. More precisely, consider the individual strategy in which a player defects against out-group members but plays nice against in-group members:

Definition 3.2 (**Network Out-Group Deviance** $NWOUT$)**.** *A player defects in*

[10] See Chapter 2 for the precise conditions under which the strategy profile forms a fully cooperative equilibrium.

all out-group pairings and plays $NWTFT$ in all in-group pairings.

Any player playing according to this strategy faces punishment from his fellow in-group members for defecting against the out-group.

The present interest is in "splitting equilibria" in which some cooperate while some cheat. Such an equilibrium depends on the following claim:

Claim 3.1. *Some players may profit more from playing $NWOUT$ than from playing $NWTFT$ and some players may profit more from playing $NWTFT$ than from playing $NWOUT$.*

To determine which players benefit more from which, we must determine the expected costs and benefits that each player faces from punishment, which depends on each player's probability of punishment.

It turns out (shown below) that the relevant punishment probabilities pertain to punishment in a particular round (the last round of the punishment phase), so Proposition 3.1 presents the probability that punishment occurs in a particular round after a deviation in t.

Proposition 3.1 (Punishment Probabilities for Out-Group Defections). *The probability that i will be punished in $t+l$ by an in-group member for defecting against an out-group member in t is*

$$z_{i,l}^{out} = \frac{\#N(i, rl)}{n-1}$$

where $N(i, rl)$ is the set of players reachable from i in rl degrees (i's neighborhood of size rl).

Now consider an ordering of players by their probability of punishment in T^p for a defection against the out-group. Call the player with the smallest value the "most peripheral" player and the player with the largest value the "least peripheral" or "most central" player. The ordering from smallest to largest ranks players from more peripheral to less peripheral/more central.

Now consider a strategy profile in which some players play $NWOUT$ and some play $NWTFT$. Let $\bar{z}^{out}$ be some cutoff value.

Definition 3.3 (Inter-Group Cheating σ^{NWOUT}). *All players with $z^{out}_{i,rT^p} \leq \bar{z}^{out}$ (the most peripheral ones) play $NWOUT$. All players with $z^{out}_{i,rT^p} > \bar{z}^{out}$ play $NWTFT$.*

According to σ^{NWOUT}, the most peripheral players will defect against the out-group when given the opportunity, and the rest will punish them for it. Clearly the profile can only form an equilibrium if the players defecting against the out-group, here the "peripheral" players, gain more from continuing to defect against the out-group than they lose in punishment from their own group. But, they also must stand to lose more in punishment were they to defect against their own group members than they would stand to gain by defecting. So, they must be peripheral enough to get away with out-group defections, and there must be a great enough difference in isolation that in-group defections would face much more dire consequences. The precise conditions under which inter-group cheating occurs in equilibrium are explored in the next section.

Inter-Group Cheating

Moving toward out-group cheating in equilibrium, consider the most peripheral players. These players defect against the out-group and face punishment for this. Whether defecting is worthwhile depends on how likely they are to face punishment for out-group defections. To ensure that defecting against an out-group member is profitable in any scenario, we must ensure it is profitable for any feasible composition of his in-group in terms of cooperators or defectors and for his knowledge of any feasible composition. The precise statement of the non-slack conditions can be found in the Appendix, but here it suffices to consider simpler sufficient conditions.

Proposition 3.2 (Inter-Group Cheating). *The individual strategy $NWOUT$ is sequentially rational for player i with $z^{out}_{i,rT^p} \leq \bar{z}^{out}$ if*

$$(\max\{\alpha, \beta+1\}) \sum_{l=1}^{rT^p} \delta^l z^{out}_{i,l} \;\leq\; \frac{\alpha-1}{1-p} \;\leq\; (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{i,rT^p} \tag{3.1}$$

and

$$\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{i,rT^p} \tag{3.2}$$

$NWTFT$ is sequentially rational for j with $z^{out}_{j,rT^p} > \bar{z}^{out}$ if

$$\frac{\alpha-1}{1-p} \leq (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{out}_{j,rT^p} \tag{3.3}$$

and

$$\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{j,rT^p}. \tag{3.4}$$

When (3.1) and (3.2) hold $\forall i$ with $z^{out}_{i,rT^p} \leq \bar{z}^{out}$ and (3.3) and (3.4) hold $\forall j$ with $z^{out}_{j,rT^p} > \bar{z}^{out}$, σ^{NWOUT} is sequentially rational.

Here T^p is the length of the punishment phase, r is the speed of information transmission, and $z_{i,l}^{out}$ ($z_{i,l}^{in}$) is the probability that i will be punished by an in-group member in $t+l$ for a defection against the out-group (in-group) in t. Conditions (3.2), (3.3) and (3.4) consider only the probability of punishment for a defection at the end of a punishment phase, $t+T^p$. This is because for these conditions, the binding scenario is one in which the peripheral or central member has formerly defected and is considering a second defection, which would earn them an additional round at the end of the punishment phase. In contrast, condition (3.1) sums over the entire punishment phase, $l = 1 \ldots T^p$, because the binding case here is a peripheral player who could forego the entire string of punishment if he opts not to defect against the out-group.

For out-group cheating to persist in equilibrium, all peripheral players must face little enough punishment when defecting against the out-group but enough when defecting against the in-group, and central players must face enough punishment in any type of defection. Take a simple example. Let $\delta = .95$, $T^p = 2$, $r = 1$, $p = .2$, $\alpha = 1.5$, $\beta = 1.05$, and let the social structure be as shown in Figure 3.1 with 21 players, 19 of which are connected to all other 18 in a clique.

Player 1's probability of being punished by an in-group member in $t+1$ for defecting against an out-group member in t, $z_{1,1}^{out}$, is .05, and in $t+2$, $z_{1,2}^{out}$, is .1. The probability that 1 is punished by an in-group member for his most tempting in-group defection (which here is any in-group defection) at the end of the punishment phase, $z_{1,2}^{in}$, is 1.

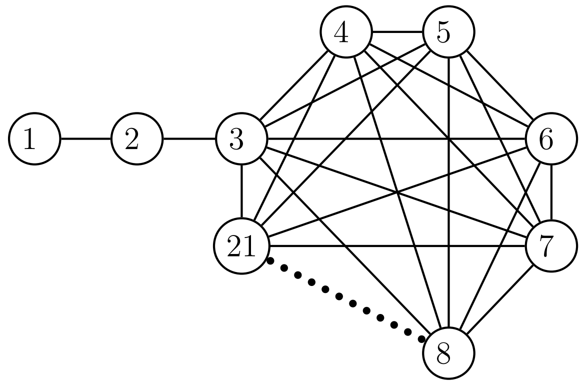

Figure 3.1: Example communication network with 21 players, 19 of which are in a fully connected clique. There exists an equilibrium in which 1 defects in out-group pairings but all in-group interactions are cooperative.

Intra-Group Cheating

When a player defects against a fellow in-group member, information spreads from his own neighbors and his victim's neighbors. That means the probability of punishment for an in-group defection is weakly larger than the probability of punishment for an out-group defection (in which information only spreads from the offender's neighbors to anyone in the group). It turns out that the smallest in-group punishment probability is the most relevant, presented in the following proposition:

Proposition 3.3 (**Punishment Probabilities for In-Group Defections).** *The smallest probability that i will be punished by an in-group member for defecting against an in-group member j in t is*

$$z_{i,l}^{in} = \min_{j}\{\frac{\#\big(N(i,rl) \cup N(j,rl)\big)}{n-1}\}.$$

Since news about in-group defections spreads from two sources, the offender and the victim, the probability of punishment for an in-group defection depends on who an offender is paired with. $z_{i,l}^{in}$ is calculated assuming i plays his most tempting opponent, the j which minimizes the chances that his defection is known.

Clearly the number of possible punishers is weakly larger for in-group defections. It is also clear that the people who punish someone for an out-group offense are also in the set of people who will punish for an in-group offense. However, now there is a cutoff in terms of the out-group offense $\bar{z}^{out}$ below which only some will do best by playing the defection strategy, but not necessarily all.[11]

When peripheral players are so isolated that defections against fellow isolated in-group members are also not widely known, intra-group cheating can occur. The following individual strategy has some players defecting against the out-group *and* against the most isolated in-group members. Now the possible defectors are those with a small enough probability of punishment by central members for in-group defections against a central member k, $z^{in}_{i,k,rT^p}|_C$ (which implies a small enough z^{out}_{i,rT^p}.

Definition 3.4 (Network Intra-Group Deviance against the Periphery $NWPER$). *All players i defect with all out-group opponents and in-group opponents k when $z^{in}_{i,k,rT^p}|_C \leq \bar{z}^{in}$ is satisfied. All other players play $NWTFT$ in all interactions.*

Under some conditions, out-group and intra-periphery defections persist in equilibrium:

Proposition 3.4 (Intra-Group Cheating). *The individual strategy $NWPER$ is sequentially rational for player i if, when playing any in-group j who satisfies $z^{in}_{i,j,rT^p}|_C \leq \bar{z}^{in}$,*

[11] This is because the ordering implied by z^{out} is not identical to the ordering implied by z^{in}; we could construct an example in which i faces a lower probability of punishment than j does for defecting against the out-group but a higher probability of punishment than j for defecting against the in-group.

$$\frac{\beta}{1-p} \geq (\max\{\alpha, \beta+1\}) \sum_{l=1}^{rT^p} \delta^l z_{i,j,l}^{in}|_C \tag{3.5}$$

and if, when playing any other in-group member k,

$$(\max\{\alpha, \beta+1\}) \sum_{l=1}^{rT^p} \delta^l z_{i,l}^{out} \;\leq\; \frac{\alpha-1}{1-p} \;\leq\; (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{i,k,rT^p}^{in}|_C \tag{3.6}$$

and

$$\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{i,k,rT^p}^{in}|_C \tag{3.7}$$

where $z_{i,k,rT^p}^{in}|_C$ is the probability that i will be punished in $t+T^p$ by a more central in-group member playing NWTFT for a defection against in-group member k in t. NWTFT is sequentially rational for j with $z_{j,k,rT^p}^{in}|_C > \bar{z}^{in}$ $\forall$ opponents k if

$$\frac{\alpha-1}{1-p} \leq (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{j,rT^p}^{out} \tag{3.8}$$

and

$$\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{j,rT^p}^{in} \tag{3.9}$$

When (3.5), (3.6) and (3.7) hold for the NWPER players (3.8) and (3.9) hold for the NWTFT players, σ^{NWPER} is sequentially rational.

In words, some players will defect against the out-group and the most isolated in-group members so long as the probability that they are punished for doing so is low enough but the probability of punishment for defecting against anyone else is high enough. Since peripheral members will defect against each other regardless of their behavior, the only relevant punishers to them are the central players. Peripheral players play nice with central players which means central players could be punished by any in-group members, central or not.

The following stylized example exhibits in-group and out-group cheating. Let $\delta = .95$, $T^p = 2$, $r = 1$, $p = .2$, $\alpha = 1.5$, $\beta = 1.05$, and let the social structure be as shown in Figure 3.2 with 21 players, this time 18 of which are connected to all other 17 in a clique.

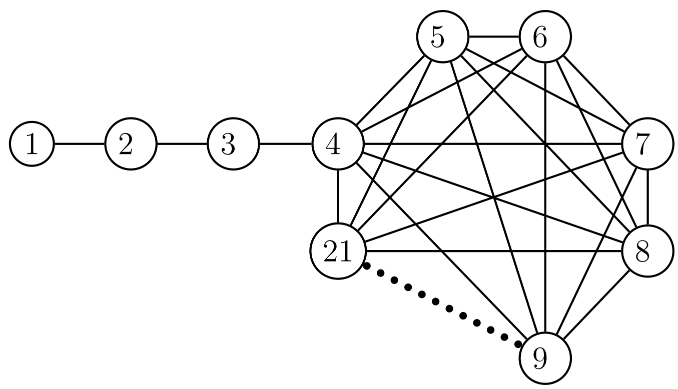

Figure 3.2: Example communication network with 21 players, 18 of which are in a fully connected clique. There exists an equilibrium in which 1 and 2 defect against the out-group and each other.

It can be verified that players 1 and 2 are isolated enough that in equilibrium they both defect against the out-group and defect against each other but cooperate with central in-group members. Player 3, while not as well connected as players in the clique, is less isolated and plays $NWTFT$ like the more central players do. Player 1's probability of being punished by an in-group member in $t + 1$ for defecting against an out-group member in t, $z^{out}_{1,1}$, is 0, and in $t + 2$, $z^{out}_{1,2}$ is .0526 since one of the 19 central players will know about the offense by then. For player 2, these probabilities are $z^{out}_{2,1} = .0526$ and $z^{out}_{2,2} = .105$. The probability that 1 and 2 face punishment for defecting against each other in the last round of the punishment phase, here $t + 2$, is

$z_{1,2}^{in} = z_{2,2}^{in} = .105$.

Note that in the above example if the punishment phase were shorter, say $T^p = 1$, the set of defecting players would be larger, containing 1, 2 and 3. If the punishment phase were longer, say $T^p = 3$, the set of defecting players would be smaller: only 1, and 1 would only defect against the out-group. By the end of a punishment phase of length 3, every central player would be aware of any in-group offense so $z_{1,3}^{in} = 1$. If the punishment phase increases still to $T^p = 4$ or longer, the probability that the most peripheral player, 1, will be punished by an in-group member at the end of a punishment phase for *any* defection is 1. This result holds in general:

Lemma 3.1 (**Punishment Duration**). *For sufficiently patient players, increasing the length of the punishment phase T^p reduces the plausibility of equilibria containing defections.*

Lemma 1 follows immediately from Propositions 3.2 and 3.4. The threat of a protracted punishment makes cheating costlier and so decreases the incentives to do so. Colson (1974, ch. 3) reviews ethnographic accounts of the Lakalai off of New Guinea, the Tonga, the Chimbu, the Pomo, the Eskimos and the Iroquois and notes agreement that that fear of protracted reprisal for misbehavior keeps these societies relatively peaceful. While Eskimo society is often regarded as violent with minimal governance (Colson, 1974, p. 41), "it was not so much perhaps the presence of feuds ... as it was the fear of feud" that generated what cooperation there was (Spencer, 1972, p.111). Likewise for the Navaho, also with "minimal government" (Colson, 1974, p.41): "[a] society like the Navaho *must* have harsh sanctions against physical aggression" (Kluckhohn, 1944, p.53). The importance of a sufficiently long and harsh punishment in order to keep everyone in line has been widely recognized.

Efficient Discrimination

While it is true that an increase in the length of the common punishment phase can reduce defections, the longer punishment would be overkill for many players. Take the example in Figure 3.2 again. A punishment of length 4 will even keep player 1 cooperating with everyone. However, player 5 would be enticed to cooperate with everyone if the punishment phase were as short as one round. In equilibrium, the observed outcome when the punishment phase is 4 and 24 is the same, but in the presence of trembles or errors, it is very different. Shorter punishment is more efficient in the presence of such deviations and excessive punishment is inefficient. Consequently, punishment too long for some players is a source of inefficiency. Akin to perfect price discrimination which avoids the waste of some paying less than their marginal value of a good, perfect punishment discrimination avoids the waste of threatening more value-destroying punishment than necessary.

For each player, if he can be enticed to be cooperative with everyone, there is a minimum length of punishment phase required to do so. The following definition identifies this minimum:

Definition 3.5 (Minimum Required Punishment). *For each player i, if $\exists$ a T^p such that* $\frac{\alpha-1}{1-p} \leq (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{i,k,rT^p}|_C$ *and* $\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{i,k,rT^p}|_C$*, then $\exists$ value T^p_i for player i such that for any $T^p < T^p_i$,* $\frac{\alpha-1}{1-p} > (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{i,k,rT^p}|_C$ *or* $\frac{\beta}{1-p} > (\min\{\alpha, \beta+1\})\delta^{rT^p} z^{in}_{i,k,rT^p}|_C$*. Call T^p_i i's **Minimum Required Punishment** (MRP_i).*

In other words, if it is possible to entice a player to cooperate with both out-group and in-group members, then there exists a smallest number of punishment rounds that will entice him to do so. This shortest punishment phase is a player's MRP. Clearly a player's MRP is larger when his probability of punishment is smaller. One solution to intra- or inter-group cheating would be to coordinate a common punishment phase that is long enough to entice the most peripheral player to cooperate:

Lemma 3.2 (**Overkill Punishment to Restore Cooperation**). *Given inter- or intra-group cheating, so long as players are patient enough, full cooperation can be attained by increasing the common punishment phase to a length at least as long as the greatest MRP. That is, for sufficiently high δ, players prefer to cooperate so long as $T^p_{new} \geq \max\{MRP_i\}$.*

Corollary 3.1 (**Inegalitarian Punishment**). *Any punishment regime which assigns to each player i a punishment phase of length at least MRP_i ensures full cooperation.*

Full cooperation is efficient in the sense that the highest total value is attained in equilibrium. However, different cooperative equilibria can be compared with respect to their efficiency out-of-equilibrium to account for the consequences of trembles or errors. One fully cooperative equilibrium is more efficient than another if its threatened punishment is more efficient (destroys less value) than the other's. Here, the most efficient cooperative equilibrium is one that maintains cooperation using the shortest possible punishment phase. By the same logic as price discrimination, with heterogeneous MRPs, the most efficient cooperative equilibrium is one which perfectly "punishment discriminates" and threatens each i with his MRP and no more.

Proposition 3.5 (The Efficiency of Inegalitarian Punishment). *The most efficient punishment regime is one in which each player i is punished according to his MRP_i. Heterogeneity in peripheral-ness implies heterogeneity in punishment lengths in the most efficient equilibrium.*

Consider once again the example group in Figure 3.3, reproduced here to indicate the MRP for each player assuming the same values of the other parameters from above.

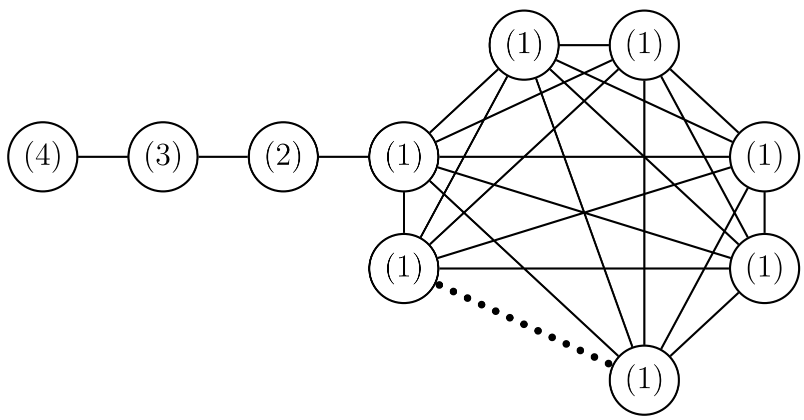

Figure 3.3: Example communication network with 21 players, 18 of which are in a fully connected clique. Values are each player's MRP assuming $\delta = .95$, $T^p = 2$, $r = 1$, $p = .2$, $\alpha = 1.5$, and $\beta = 1.05$.

The most peripheral player requires the longest punishment to be kept in line, the next-most-peripheral player requires the next longest, and so on. Of course, calculating each player's exact MRP and keeping track of who should be punished for how long may be taxing in practice. We might expect punishment regimes that treat subsets of the in-group differently, making coarse groupings to avoid complicated, tailored punishment regimes. So long as such regimes threaten sufficient punishment to each group, they can maintain cooperation more efficiently than a regime with uniform punishment.

For example, the group shown in Figures 3.2 and 3.3 might adopt a differentiated punishment regime instructing players to punish the three most peripheral players for 4 rounds but to punish the rest for a single round.

We might expect real world networks like the one in Figures 3.2 and 3.3 to exist in places with a stark difference between the town and country populations, or in places in which a few new people just moved to town. The above reasoning suggests that groups might do well to assume that all newcomers will be more troublesome than they really are. By assuming that the new population is as prone to cheating as the worst exemplar, it can maintain cooperation. So long as the old population applies a shorter punishment phase to themselves, the group can maintain cooperation more efficiently than if they treated everyone as deserving the same punishment for the same offense. The discrimination of newcomers may have its origins in an attempt to balance efficiency and manageability.

Taking this logic a step further, protracted feuds can be understood as one solution to a problem of group heterogeneity. When some players face much greater incentives to defect initially or are much more isolated than the rest of the group, threatening long-lasting feuds makes sense. If the length of the feud was miscalculated or something happens to pull the group out of equilibrium, the Hatfields and McCoys are born. More on this below.

The Dangers of Clustering

The structure of the peripheral players bears on whether they are likely to defect only against the out-group, or against both the out-group and the in-group. Typically there is thought to be value in tight-knit groups, especially in the context of violence or crime. The more cohesive a group, the better a neighborhood watch can function. The following proposition suggests that this view is overly simplistic.

Proposition 3.6 (**Dangerous Clustering**). *All else equal, the more clustered*[12] *the peripheral members, the more likely they are to defect not just against the out-group, but against each other as well.*

The proof of Proposition 5 can be found in the Appendix, but for the intuition, consider the following example in Figure 3.4.

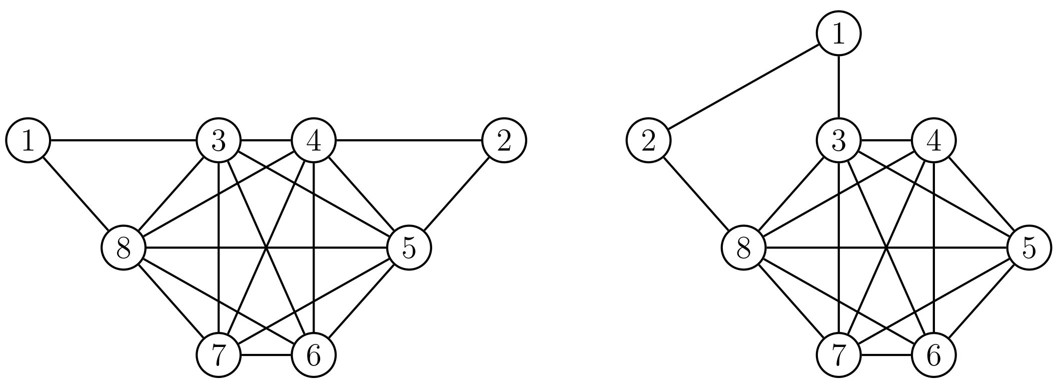

Figure 3.4: Two example communication networks, each with 8 players.

In both networks, the two most peripheral players, 1 and 2 have the same degree, each with 2. For simplicity, if $r = 1$ and $T^p = 1$, then the probability of punishment that either faces for defecting against the out-group in either network is $\frac{2}{7}$.[13] The probability of punishment for defection against a fellow peripheral member is different in the two networks, though. In the left network, if 1 and 2 defect against each

[12]Clustering is a measure of the extent to which neighbors of nodes are themselves neighbors; i.e. of the prevalence of closed triples.

[13]or $\frac{2}{6}$ if they are also playing *ALLD* against each other. The point is they face the same punishment probability.

other, the neighbors of both observe, so 5 of the 7 other players could punish. In the right network, if 1 and 2 defect against each other, the neighbors of both observe, which this time only results in 3 of the 7 other players.[14] Clustering among peripheral players shrinks the difference between a player's probability of being punished for defecting against the out-group and a player's probability of being punished for defecting against the in-group.

A more clustered periphery can better shield information of their defections from the rest of the group, most importantly from the central members who could punish them. Rather than serve as a neighborhood watch, when clustering occurs among those who have an incentive to do something wrong anyway, it insulates against watch by the more pious members of a group.

Thanks to information about residency patterns, we have some idea of the clustering of some peripheral subgroups in frontier boomtowns, discussed below.

Intra-Group Bullying and Raiding

The cheating above occurs under the assumption that information spreads from both the victim and the offender, and spreads equally quickly and well. Not only does this simplify matters, it comports with the idea of neighbors as observers. If the people player i talks to see what i is doing, then whether i is doing something right or doing something wrong, they know about it. This assumption is most likely to hold for actions occurring within reasonably close proximity, like a marketplace, or for significant, visible actions like extreme violence. And of course it is possible that

[14] If we stipulate that the center 6 players play $NWTFT$ and the others play $NWPER$, there are 6 total possible punishers, 4 of which know in the left network, 2 of which do in the right one.

people are by nature bad at keeping secrets and spill news, even about themselves, even if the news sheds negative light on themselves.

Surely there are some instances, though, in which actions are not as visible to everyone or in which some people fail to reveal information. Suppose that actions are hidden, and that only victims tell their neighbors about an offense and start the chain of information spread.[15] Clearly now the probability of punishment for an in-group defection is no longer determined by the union of players reachable from both the victim and the offender; it is determined only by the players reachable from the victim. In other words,

Lemma 3.3 (Punishment Probability for Asymmetric Info from the Victim). *When news of a defection spreads only via the victim, the probability that i will be punished in $t+l$ for a defection against an in-group member j in t is*

$$z^{in}_{i,j,l} = \frac{\#(j, rl)}{n-1}$$

where $\#N(j, rl)$ is the number of players reachable from the victim j in rl degrees.

Note that the probability is determined solely by the identity of the victim. This means that regardless of the position of a prospective offender, the probability of defecting against any particular in-group member j is common to all. In other words,

Corollary 3.2 (Piling On). *When news of defections spreads only from the victim, if anyone has an incentive to defect against a particular in-group member, all have an incentive to defect against that in-group member.*

[15] Suppose information is still not being manipulated strategically. Modify the information process so that all people pass along only what they play, and only the victims send an additional message indicating they were cheated, all done truthfully and deterministically. Strategic manipulation is explored in Chapter 4.

This scenario can be considered bullying, and can be most clearly seen in a very stylized example of a disconnected in-group like the one in Figure 3.5.

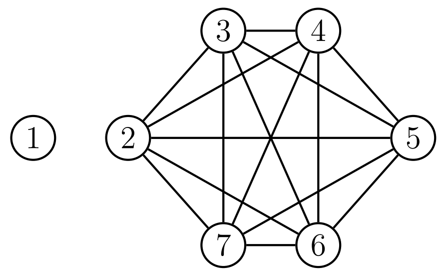

Figure 3.5: Example communication network with 7 players, one of which is extremely peripheral.

Again, by assumption, only the victim spreads information about an offense. Clearly here, 1 is a very tempting victim. Any of the other players could defect against him, and since he has no one to tell, no one will find out. In this extreme case, 1 serves as the punching bag for the rest of the bullies who face no punishment so long as they can keep defecting on 1 in secret.

Consider the opposite scenario. Suppose defections take a form such that only the neighbors of offenders observe them. This is likely to be the case when the act of defecting is itself secret, but only neighbors of the offender observe the consequences.[16] Take a cattle rustler who robs cattle from a huge herd in the middle of the night. The next day the rancher doesn't notice cattle are missing, but the neighbors of the rustler see new misbranded cattle in his smaller herd. In such a scenario, the probability of punishment is determined solely by the prospective offender.

Lemma 3.4 (Punishment Probability for Asymmetric Information from the

[16] This is also likely to be the case when there is taboo or shame to being victimized. If the offense is embarrassing or shameful, we might expect peripheral stalker types targeting more central group members with impunity since taboo keeps them from discussing the offense.

Offender). *When news of a defection spreads only via the offender, the probability that i will be punished in $t+l$ for a defection against an in-group member j in t is*

$$z^{in}_{i,j,l} = \frac{\#(i, rl)}{n-1}$$

where $\#N(i, rl)$ is the number of players reachable from the offender i in rl degrees.

In this case, the opposite corollary obtains:

Corollary 3.3 (Company in Misery). *When news of defections spreads only from the offender, if someone has an incentive to defect against a particular in-group member, he has an incentive to defect against all in-group members.*

If the group is again like Figure 3.5, this time player 1 is *tempted.* Since news of any of his offenses would travel only through him, and he talks to/ is observed by no one, he is a candidate to raid any of the other in-group members. In other words, 1 is the most likely cattle rustler of the group.

Medieval Iceland had little in the way of formal governing apparatus and the population was largely dispersed: "[t]he basic unit of residence, production, and reproduction in Iceland was the farmstead. Until the end of the eighteenth century, there were no villages or towns, no nucleated settlements at all" (Miller, 1990, p.15). The organization of daily life suggests few strong ties and conjures an image of medieval farmers existing in nearly complete isolation from all others. "The landscape, providing no trees for cover, the settlement pattern, with each farm situated independently of others, made hostile approaches self-evidently visible" (Miller, 1990, p.188). Such a landscape matters for hostile approaches and also for cohesion with near neighbors. In such a setting, the isolated (and hence peripheral farmers) were occasionally bullied by others, as when chieftain would raid the farmer to provision an army (p. 185). The punishment the isolated farmer could wage on the chieftain was minimal at best, and only occasionally could he get the word out widely enough or commission another chieftain in order to levy punishment. With such a small punishment threat coming from the peripheral farmers, it is no wonder they were tempting targets.

Dangerous Outskirts

The American frontier in the 19th century turns out to be an ideal setting for examining the plausibility of the above results. While it has become fashionable to regard the 'Wild West' as "a far more civilized, more peaceful and safer place than American society is today," (Hollon and Crowe, 1974, p.x), formal governing institutions were largely absent and weak when present. In fact, revised histories of the frontier usually start with the observation that the West was surprisingly cooperative and peaceful despite the poor formal governing institutions (see, for example, Prassel, 1972; Anderson and Hill, 2004).[17]

Even though more and more evidence suggests that the settlers on the frontier did not face as much danger and chaos as had been earlier assumed (see McGrath, 1987), the environment is a natural setting for examining cheating for two reasons. One, because the formal governing institutions were often weak at best, groups of individuals were left to develop informal means of getting along. Two, records of frontier violence and treachery are unusually detailed for a weak government setting.

[17]Boom towns were especially prone to ineffective governing institutions, if only because the growth of the towns outpaced formal governance (and sometimes the construction of a jail). Halaas (1981) notes that "in those mining districts where legally constituted law enforcement agencies were either ineffective or nonexistent, editors encouraged the law-abiding population to use extralegal means of quieting chronic lawbreakers and violators of the public peace" (p.85).

Such detail is especially important for a micro-oriented approach like the present one, in which the identity of offenders matters.

Thanks to the work of McGrath (1987), such a detailed account is available for the mining town of Aurora in modern-day Nevada in the Sierra Nevada. Aurora was a classic boom town, growing from a few prospectors striking luck in the mines to thousands of residents in the course of a year (McGrath, 1987, p. 9). In the height of the boom, with 5,000 residents, "Aurora was bursting at the seams. Every hotel, lodging house, and miner's cabin was jam-packed, and hundreds of people went without accommodations" (p.9). The miners quickly outstripped supplies, and nearby cattle ranchers capitalized on the new market by driving cattle in and setting up ranches in the hills around the town (p.17). The settlers then occupied a very dense town center and a sparser set of ranches in the hills.

Of course, none of the land was actually previously unused, and relations with Paiute who relied on the land for foraging and growing roots grew tense. After some initial conflict between the white settlers and the Paiute, both parties met and drew up an informal treaty to promote inter-group peace (p.20). If the framework presented here applies, we should see white settlers trying to keep other white settlers from violating the treaty, and likewise for the Paiute. We should also expect that if cheaters (treaty violators) exist, they should be the more isolated group members, and that if some shock occurs which uniformly increases incentives to cheat, the most isolated group members should be the first to cheat.

Records are more detailed for the white settlers than for the Paiute, but what we see is consistent with the predictions. It does seem to be the case that in-group members tried to police their own (p. 49); despite the scarcity of evidence from within the Paiute, there are even records of Paiute counseling their own against war (p. 23). The winter following the treaty, of 1862-63, was especially severe and created much greater scarcity of the resources supposed to be shared by the two groups (p. 20). Such a shock increased the incentives of everyone to steal from others and violate the treaty. The first instance of inter-group violence in violation of the treaty was by a Paiute known as "Joaquin Jim." "Joaquin Jim" was an outcast among his people (p.21), which surely counts as peripheral. Among the white settlers, those who first violated the treaty were the more isolated ranchers and not the more densely packed prospectors in town (p. 18).[18] Those most isolated and peripheral were the more difficult to convince to respect the treaty and the most likely to commit inter-group cheating.

In a nearby mining town–Brody–a bit more information is available about the internal composition of the miners. Brody, also experiencing a population boom driven by hopes of striking it rich in the mines, had relatively sizable Chinese and Mexican minorities. The Chinese carved out a separate part of town at the far northern edge; the Mexican miners were scattered throughout the town (p. 140). Few learned English or desired to assimilate (p.124), so presumably few forged social ties with the other miners. Once again, the patterns of cheating, here violent, anti-social behavior, are consistent with what we would expect given the immigrants' likely position in the miners' communication network. Both the Chinese and Mexican

[18] Also of note is that violence persisted as a string of isolated incidents through the fall and winter of 1863. Small-scale conflict can be durable, even if not permanent. Joaquin Jim acquired a band of followers who "remained at large and continued to prey on the unsuspecting traveler or prospector" (p. 40).

miners were likely peripheral subgroups in the community of miners, and indeed there was disproportionate in-group violence within these peripheral groups (p.133, 142). Since the Chinese lived in a Chinatown but the Mexican miners were dispersed throughout, it is likely that the peripheral Chinese were more connected to their fellow Chinese (i.e. they were more highly clustered within the periphery). From above, we would expect the clustered group of Chinese miners to pose a greater threat to non-Chinese miners than the unclustered group of Mexican miners would. While the historical record is not rich enough to offer unambiguous differences, there are accounts of a few violent encounters between Chinese and white miners outside of opium dens, and plenty more inside opium dens (p.133, 129) and fewer accounts of violence between Mexican and white miners (p.140).

Finally, the above suggests that when there are sufficient differences between how easily people can be kept in line, we should expect pressure toward differentiated punishment, and possibly even pressure to exclude some from the in-group. Indeed, Brody faced pressure to expel the Chinese. Such an exclusionary view was not uncommon in 1870s California, and we may be tempted to attribute it strictly to the wave of racism sweeping the frontier. On the other hand, such a view may have been at least reinforced by the perception, however unjust, that the Chinese miners must be subject to harsher punishment to keep peace. Putting the issue harshly, the *Daily Free Press* printed in February 1880: "We reflect the sentiment of a large majority of the citizens of this coast when we say that we have no desire to see the Chinese ill-used or badly-treated in any way, but they are a curse to the people of the coast, and we do not want them here. They do not and cannot assimilate with Americans..."

(McGrath, 1987, p.137). However unsavory, the above suggests that when, due to their position in the communication structure, some are more difficult to keep from cheating, and when deeper assimilation into the group fails, we should expect a push toward exclusion.

Feuds Between Those Farther Apart

Reexamining the work of Evans-Pritchard (1940), Gluckman (1955) marvels at the stability and complexity of relations among the Nuer. Despite the lack of formal governing institutions, the Nuer abide by an established code of conduct that covers a range of social, political and economic behavior and is remarkably successful at securing peace. Gluckman attributes this success to the overlapping ties that arise from rules of exogamy and relationships forged due to the seasonal mobility of the group (relocating from low to high ground during the flood season and back to low during the dry season). Such ties produce "sufficient conflicts of loyalty" (p. 10) which foster peace. Of course peace is not ubiquitous, and Gluckman notes "I do not wish to give the impression that vengeance is never taken and the feud never waged. *Feud is waged and vengeance taken when the parties live sufficiently far apart, or are too weakly related by several ties.*" (p.11, emphasis added).

The model above examines the role of ties that transmit information from person to person in mitigating the incentives to cheat, or in the case of the Nuer, engaging in feuding. To understand the role that distance might play in incentives to feud or not, it is important to consider the relationship between distance and communication. It is reasonable to assume that, even though the Nuer are seasonally mobile, social contacts correlate with geographic distance. Indeed, Gluckman observes: "[t]he ecological needs for this friendship and peace lessen as the distance grows greater, until, between districts on opposite sides of a tribe, it hardly exists" (p.4). Similarly, Evans-Pritchard (1940) remarks that "[i]n a small group like the village not only are there daily residential contacts, often of a co-operative nature, but the members are united by close agnatic, cognatic, and affinial ties which can be expressed in reciprocal action. These become fewer and more distant the wider the group..." (p. 138). While it could be that those farthest apart also so happen to be the least integrated into the group so that these are the problem peripheral players identified above, consider the harder case for the importance of distance in which all players are as connected as all others (figure 3.6).

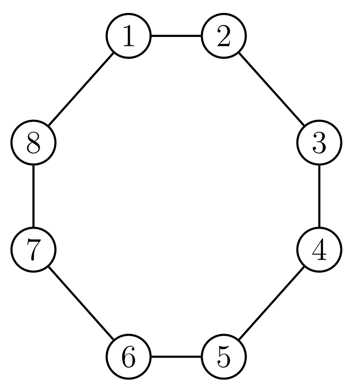

Figure 3.6: Example communication network with 8 players, where node placement is geographically meaningful.

Imagine that figure 3.6 is not just a representation of the communication links between people, but also indicates who lives where. In other words, the placement of the nodes is geographically meaningful, and geographic distance can be represented with geodesic distance.[19] Here, each player is geographically closest to his neighbors

[19]In other words, the more links that separate two nodes, the more geographic distance between

in the network, and farthest from the person four links away (diagonally across from the player).

If players were matched uniformly at random and interacted in a marketplace where neighbors of both players could observe them (as in the setup above), then player 1 has no greater incentive to feud with player 5 (the person farthest away) than with player 6 or 4. Greater geographic distance, even if it means greater social distance, is insufficient to produce more mutual cheating.

Actually, in this simple network, player 1 has a greater incentive to feud with his neighbors, 8 and 2, than with anyone else. As in the case of clustering above, the total number of group members who can observe 1's actions when playing 8 are 3 (8 himself, 7 and 2). If player 1 would play 5, the total number observing who could spread the word and would later punish are 5 (player 5 himself, 6, 4, 8, and 2). When all players are equally connected, defecting against someone nearby is easiest to get away with, since the offender is one of the people who would usually have the victim's back, and the victim is someone who would usually keep the offender in line. Clearly, if there are not differences in connectivity between those nearby and those far apart, this system of meeting and spreading news will not make feuds between those far apart the most likely feuds.

Consider two modifications to the setup:

1. Players encounter each other at one or the other's home

2. Players are more likely to encounter their neighbors than anyone else

Modification 1 is a change so that players no longer meet in an easily observable

the two.

marketplace, but instead encounter each other near one of their homes. A player may have occasion to pass the other's property regularly. When players in the game are paired, suppose that the roster not only declares who plays whom, but also on whose turf, also selected randomly.[20] Suppose interactions on someone's turf are observable to the neighbors of that area. In other words, if 1 visits 4, their interaction is observable to 4's neighbors but not 1's. This modification alone is not enough to favor long-distance feuds; it simply attenuates the preference for defecting against a neighbor over someone else. The second modification makes a player more likely to encounter his own neighbors than anyone else. In other words, players are matched according to a weighted probability that favors matches between neighbors.

When a player is more likely to encounter a neighbor, defecting against a neighbor is especially costly since he can punish and will have a greater opportunity to do so. Under the two modifications, a player's most profitable cheat would be to defect against the person farthest away while visiting that person. Only neighbors of the person far away observe the action, and the offender is most likely to play his own neighbors who are kept in the dark for as long as possible.

In short, when the social structure is related to geography so that social distance is correlated with geographic distance, all else equal, players have incentives to feud in close proximity when they encounter others on common, observable ground uniformly at random, and incentives to feud with those far away when they encounter others on personal turfs and encounter their own neighbors most frequently.

[20] Think of something akin to a roster of football matches which also declares one team in each match to be the home team.

Discrimination and Duration

The above identifies cheating that may occur due to heterogeneity in the extent to which news about individuals' actions spreads. Cheating can occur by peripheral members against the out-group and can even occur within the periphery, with the latter more likely if clustering is higher among the periphery. Section 5 establishes that discriminatory punishment can mitigate cheating of either form so long as players are patient enough. Here I want to delve deeper into the issues of discriminatory punishment and the durability of these equilibria.

First, about discriminatory punishment. Groups with problem peripheries would do well to threaten longer punishments for peripheral members acting out of line. This logic extends naturally and dangerously. It is not a difference in the kind of actions taken by peripheral and non-peripheral members, but a difference in incentives both face. Both could take the same misstep but be punished differently for it. Not only could this help to shed light on persistent feuds that break out and last longer than any similar conflict between different members of the same community, but could also suggest a source of pressure for assimilation or segregation.

An option for sparing central members the risk of peripheral conflict is to excise the periphery. While the above game does not allow manipulations of the communication network or changes in the frequency of interactions, surely a violent periphery creates pressure to reshape interactions to make their violence inconsequential. One such option is to cease interactions with the periphery, to regard the periphery as "others" who are no longer members of the group.

Another such option is to make the periphery less peripheral. Violence in the periphery is less likely when the violent members are observed more closely. Drawing them in to the primary social circle of a community is an option. Such assimilationist tactics may be behind some group's willingness to embrace newcomers warmly and quickly. Perhaps some of the variance behind the understanding of kin and how inclusive or exclusive the working definition is among groups (see, for example, Dunning and Harrison, 2010).

Second, in light of the options to improve cooperation, a fair question is why would these cheating equilibria persist? One answer is they might not; these equilibria might be temporary. Temporary does not mean trivial, however. Given evolutionary pressure toward efficiency and cooperation, the most plausible source of cheating equilibria is a change in underlying parameters and accompanying stickiness. The equilibria may not endure forever, then, but may persist for a significant amount of time, depending on how sticky the parameters. Take an example in which a small group is completely connected- all communicate with all others regularly. Then, one new person moves to the community and makes friends with one native. The punishment threat used to keep the original members in line may not suffice to keep the new person in line. Coordinating a different punishment for the new person, or a longer punishment for everyone, is a process that takes time and may be costly. Likewise, a formerly well-connected group may have weathered a natural disaster which made regular communication much more difficult with a few now poorly connected people. Violence may erupt as a consequence, and repairing the means of communication may be a long and, again, costly process.

Conclusion

When some players are poorly integrated into a group, they may know little about others and little may be known about them. Such players may face a greater incentive to cheat out-group members, fellow peripheral in-group members, or even more central in-group members. If a more central player could defect without being detected by his own neighbors, then peripheral players are the most tempting targets of cheating. In any case, the presence of peripheral players is the source of the problem.

Since new members of a population are likely to be less integrated into the social and communication network than the native population, it is not surprising that "boom towns" experience more crime than towns with stable populations. It is also not surprising that not all crime in boom towns is committed by the newcomers; the above suggests some could be committed against the newcomers too.

While the factors that improve or deteriorate cooperation are likely sticky, groups do have options for reducing the extent to which peripheral members cause problems. The group could coordinate a modified punishment regime that punishes peripheral players for a sufficient length of time. This fix poses a tradeoff though: the group could impose a punishment regime that continues to punish all players the same for the same offense, but such a punishment regime will be inefficient since it threatens more than necessary for some players. Or, the group could impose a regime that is efficient in and out of equilibrium but that is discriminatory. The incentives for discriminatory punishment are akin to the incentives to make 'others' out of the peripheral members and even to excise them from the group.

Of course these remedies have limited effectiveness in networks that are not connected and will only work if players are patient enough. If the immediate future matters much more than the distant future, cheating will persist. Tight-knit groups may improve monitoring as a whole, but tight-knit subgroups can be dangerous. When peripheral members are highly clustered, they have an increased incentive to defect against more people.

In short, communication networks determine how well a group can enforce cooperation. Some groups for which full cooperation is impossible have a next-best option of partial cooperation.

Appendix: Proofs

Proof of Proposition 3.2. Sequential Rationality and Sequential Equilibrium

σ^{NWOUT} is sequentially rational at any information set given any specification of beliefs for game G with gossip networks $\mathsf{g}_A = \mathsf{g}_B$ if, for all peripheral players i playing $NWOUT$,

$$(\max\{\alpha, \beta+1\}) \sum_{l=1}^{rT^p} \delta^l z_{i,l}^{out} \;\leq\; \frac{\alpha-1}{1-p} \;\leq\; (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{i,rT^p}^{in}$$

and

$$\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{i,rT^p}^{in}$$

and for all central players j playing $NWTFT$,

$$\frac{\alpha-1}{1-p} \leq (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{j,rT^p}^{out}$$

and

$$\frac{\beta}{1-p} \leq (\min\{\alpha, \beta+1\}) \delta^{rT^p} z_{j,rT^p}^{in}$$

where $z^{out}_{i,l}$ and $z^{in}_{i,l}$ are the probability that i will be punished in $t+l$ for a defection against an out-group member in t and for the least-punishable defection against an in-group member, respectively.[21]

Parameters z^{out} and z^{in} depend on the network g and are presented in Chapter 2.[22] Each player's strategy in each round conditions only on observable information: whether or not the player has received information that his partner is a defector.[23] However, for an incomplete gossip network g, at least one player may have incomplete information about the full history of the game. Some players will not punish some defectors since they do not know about the defections. This is unproblematic for the game, since their strategies only instruct them to punish when they know.

A second consequence of observing only a partial history of play is more subtle. To play a best response, players must compare the expected value of playing the candidate equilibrium strategy to the expected value of defecting. These expected values depend on the probability of encountering other defectors, specifically on the probability of encountering other defectors *that the player knows are defectors.* Sequential rationality requires considering this comparison for all histories. It will suffice here to consider this comparison in all states of informedness a player may be in.[24] If a

[21]The condition of symmetry between both groups' networks $\mathsf{g}_A = \mathsf{g}_B$ is an unnecessary simplification. When $\mathsf{g}_A \neq \mathsf{g}_B$, both networks must satisfy the above conditions, which means the condition binds for the least punishable player in either network.

[22]I drop the explicit dependence on g for easier reading. These parameters could be written $z(\mathsf{g})^{out}_{i,l}$ and $z(\mathsf{g})^{in}_{i,l}$. The mapping from an arbitrary network g to $z(\mathsf{g})^{out}_{i,l}$ and $z(\mathsf{g})^{in}_{i,l}$ can be found in Chapter 2.

[23]This information is observable since the network and the rules governing how information spreads are common knowledge. When a player defects, he can determine who will know in each future round. And of course, he knows who he knows about.

[24]This suffices since, in an incomplete network, the history of play may result in all players being defectors, say, but the player faces a limit to the number of defectors he could possibly know about

player only knows his state of informedness and not the true history, he can't be sure what his state of informedness will be next round, or the round after, and so on. We could specify how a player should make a best guess about their future informedness, but it will turn out that the condition that binds does not depend on this guess, so we can leave beliefs about future informativeness unspecified and show players' best responses regardless of how they form beliefs or guesses about the play they do not observe.

I will show that given proposition 1, no player has an incentive to deviate from σ^{NWOUT} for a single stage in any possible history of play and any beliefs about the history of play. Since payoffs are continuous at infinity, preventing single-stage deviations for any history and any beliefs about history ensures no deviations of any length and hence sequential rationality.[25]

The relevant global history up to time t can be fully captured by the punishment status of each player at the beginning of round t. Modifying the approach in Fearon and Laitin (1996), let vector $s_{i,t} = (t_1, t_2, \ldots, t_n)$ characterize the punishment status that i knows all n individuals deserve at the beginning of period t. Based on i's received gossip and experiences, i may not know about every defection which occurred, so i may underrepresent defectors relative to the true global history, and $s_{i,t}$ does not necessarily equal $s_{j,t}$ for player $j \neq i$. $t_{i,2} = 1$ means player i knows player 2 deserves

based on his position in the network and the speed of gossip. Another way to say this is that the mapping from histories to informedness is not unique; for some players, different histories will look the same because they will result in knowledge of the same defectors and lack of knowledge of the rest regardless if they really defected or not. Their state of informedness governs their play and others play when paired with them.

[25]This is related to the single-stage deviation principle in Fudenberg and Tirole (1991) typically used to show subgame perfection. Here the only proper subgame may be the entire game, so subgame perfection has no bite.

one more round of punishment; $t_{i,2} = T^p$ means that player i knows 2 defected in the last round and deserves punishment for T^p rounds; $t_{i,2} = 0$ means player i does not think player 2 deserves punishment.

When i must make a best guess about his future information, as when in t he must predict which (if any) defectors he will know about in $t+1$, his guess about his future information will be indicated with a $*$. As discussed, how he makes this guess will turn out to not matter, so I leave the formation of these beliefs unspecified. The thought experiment used to test sequential rationality assumes that for some arbitrary history, players switch to σ^{NWIGP}. Since players may not observe all other play, they may not know in t which defections they will know about by $t+1$ or by $t+2$ and so on until $t+T^p$. Since strategies are common knowledge and the thought experiment stipulates all switch to σ^{NWOUT} in t, by $t+T^p$ all players playing $NWTFT$ (the central players) should be cooperative. Some players playing $NWOUT$ (the peripheral players) will also be cooperative because they might not have been assigned to play an out-group member in the relevant timeframe and so are cooperative by chance of opportunity; the expected number of such players can be calculated. Players only need to make a guess about periods $t+1$ through $t+T^p-1$.

A player has a true status, and a status in the eye of his opponent. A player who has defected a number of periods less than the punishment phase ago is a true defector, a player who has not is a true cooperator. A player whose opponent knows he defected a number of periods less than the punishment phase ago is an apparent defector, and a player whose opponent does not know this (possibly because he is a true cooperator) is an apparent cooperator. We must show that no player has an incentive to deviate in any way regardless of what his opponent knows about him in any history for any guess he makes about his own future informedness. Accounting for a possibly incomplete gossip network, there are seven cases of possible deviation to consider for those playing $NWTFT$ (all central players and peripheral players playing in-group players) and seven for those playing $NWOUT$.

A player i who has not defected in the past T^p rounds (a so-called "cooperator") can deviate from $NWTFT$ by

(Ic) cooperating with an apparent defector (i.e. failing to punish someone he knows deserves it),

(IIc) defecting against an apparent cooperator,

(IIIc) defecting against an out-group member.

Likewise, a player i who has defected sometime in the past T^p rounds (a so-called "defector") can deviate from $NWTFT$ by

(Id) cooperating with an apparent defector (i.e. failing to punish when he knows punishment is deserved),

(IId) defecting against an apparent cooperator when the opponent knows i is a defector (i.e. when he is an apparent defector),

(IIId) defecting against an apparent cooperator when the opponent does not know i is a defector (i.e. when he is an apparent cooperator),

(IVd) defecting against an out-group member.

And, a cooperator can deviate from $NWOUT$ by

(Icc) cooperating with an apparent defector (i.e. failing to punish someone he knows deserves it),

(IIcc) defecting against an apparent cooperator,

(IIIcc) *cooperate* with an out-group member.

(Idd) cooperating with an apparent defector (i.e. failing to punish when he knows punishment is deserved),

(IIdd) defecting against an apparent cooperator when the opponent knows i is a defector (i.e. when he is an apparent defector),

(IIIdd) defecting against an apparent cooperator when the opponent does not know i is a defector (i.e. when he is an apparent cooperator),

(IVdd) *cooperate* with an out-group member.

Consider deviation (IVcc). A cooperator i prefers to comply and defect against the out-group if the gains to doing so outweigh the gains from deviating and cooperating with the outgroup, which obtains iff

$$\alpha - 1 \geq (1-p)\sum_{l=1}^{rT^p} \delta^l \left[prob(play\bar{C}^*)_{i,l}(\beta+1)z_{i,l}^{out}|\bar{C}^* + prob(play\bar{D}^*)_{i,l}\alpha z_{i,l}^{out}|\hat{D}^*\right]$$

and a defector i prefers to comply rather than deviate according to (IVdd) iff

$$\alpha - 1 \geq (1-p)\sum_{l=t_i}^{rT^p} \delta^l \left[prob(play\bar{C}^*)_{i,l}(\beta+1)z_{i,l}^{out}|\bar{C}^* + prob(play\bar{D}^*)_{i,l}\alpha z_{i,l}^{out}|\hat{D}^*\right]$$

Since the most challenging history is the one in which i is at the end of his punishment phase and so $t_i = 1$, the condition for preventing (IVdd) is the same is the condition for preventing (IIIcc).

As shown in more detail in Chapter 2, a player prefers to comply than deviate according to (Icc), (IIcc), (Idd), (IIdd) and (IIIdd) iff

$$\alpha - 1 \leq (1-p)\delta^{rT^p}\left[prob(play\bar{C}^*)_{i,rT^p}(\beta+1)z^{in}_{i,rT^p}|\bar{C}^* + prob(play\bar{D}^*)_{i,rT^p}\alpha z^{in}_{i,rT^p}|\hat{D}^*\right]$$

and

$$\beta \leq (1-p)\delta^{rT^p}\left[prob(play\bar{C}^*)_{i,rT^p}(\beta+1)z^{in}_{i,rT^p}|\bar{C}^* + prob(play\bar{D}^*)_{i,rT^p}\alpha z^{in}_{i,rT^p}|\hat{D}^*\right]$$

Likewise, players prefer to comply with $NWTFT$ than deviate iff

$$\alpha - 1 \leq (1-p)\delta^{rT^p}\left[prob(play\bar{C}^*)_{i,rT^p}(\beta+1)z^{out}_{i,rT^p}|\bar{C}^* + prob(play\bar{D}^*)_{i,rT^p}\alpha z^{out}_{i,rT^p}|\hat{D}^*\right]$$

and

$$\beta \leq (1-p)\delta^{rT^p}\left[prob(play\bar{C}^*)_{i,rT^p}(\beta+1)z^{in}_{i,rT^p}|\bar{C}^* + prob(play\bar{D}^*)_{i,rT^p}\alpha z^{in}_{i,rT^p}|\hat{D}^*\right]$$

Now, since $prob(play\bar{C}^*)_{i,l}z^{out}_{i,l}|\bar{C}^* + prob(play\bar{D}^*)_{i,l}z^{out}_{i,l} = z^{out}_{i,l}$ and the same for $z^{in}_{i,l}$, sufficient conditions for disincentivizing deviations from $NWOUT$ can be written:

$$\alpha - 1 \geq (1-p)\max\{\alpha, \beta+1\}\sum_{l=1}^{rT^p}\delta^l z^{out}_{i,l},$$

$$\alpha - 1 \leq (1-p)\delta^{rT^p}\min\{\alpha, \beta+1\}z^{in}_{i,rT^p},$$

and

$$\beta \leq (1-p)\delta^{rT^p} \min\{\alpha, \beta+1\} z^{in}_{i,rT^p},$$

which condense to the conditions in Proposition 3.2. Analogous sufficient conditions can be written for $NWTFT$. Therefore, when the conditions in Proposition 3.2 hold, no player has an incentive to deviate in any history given any beliefs. Therefore, the conditions are sufficient for sequential rationality.

The behavior in equilibrium is of greater interest than the beliefs in equilibrium. The above shows that σ^{NWOUT} forms half of a sequential equilibrium since the strategy is sequentially rational in any information set with any beliefs over nodes in the information set. Consider a set of beliefs, μ, which places probability one on all events expected in equilibrium, probability zero on out-of-equilbrium events, and which are updated according to Bayes' rule. Now consider a perturbation of these beliefs in which players are assumed to tremble with small independent probability $\epsilon > 0$. Given the perturbation, all information sets are reached with positive probability and Bayes' rule pins down beliefs everywhere. Let μ^* be the limiting beliefs derived from Bayes' rule as $\epsilon \rightarrow 0$. Since sequential rationality holds for any specification of beliefs given the above conditions, it holds for μ^*. Hence, the assessment (σ^{NWOUT}, μ^*) is a sequential equilibrium.

$\square$

Proof of Proposition 3.1. The proof follows immediately from the specification of information spread. Since information spreads r degrees from i each period after a defection in t, so $\#(i, rl)$ will know about the defection and be capable of punishing.

Any could be randomly assigned to i in l, making the probability

$$z_{i,l}^{out}\frac{\#(i,rl)}{n-1}$$

and the same for in-group defections, in which case news spreads from both i and opponent j. □

Proof of Proposition 3.4. Proposition 3.4 can be proven using the same technique as the proof of Proposition 3.2, with the change that now intra-group cheaters can only be punished by defections by $NWTFT$ players, which attenuates punishment. Calling punishment conditional on playing a $NWTFT$ player $z|C$, the conditions of Proposition 3.2 follow. □

Proof of Lemma 3.1. Players prefer to defect against the out group only if

$$\alpha - 1 \geq (1-p)\max\{\alpha, \beta+1\}\sum_{l=1}^{rT^p}\delta^l z_{i,l}^{out}$$

Since z^{out} is increasing in T^p, the right hand side of the condition is increasing in T^p. □

Proof of Lemma 3.2. Lemma 3.2 is immediate from Propositions 3.2 and 3.4. All players prefer to cooperate when the punishment phase is long enough, and “long enough” is determined by a player's MRP. □

Proof of Proposition 3.5. From Lemma 2, a player's MRP is sufficient for inducing cooperation. Any shorter punishment phase will be insufficient for inducing cooperation, and any longer punishment is superfluous. Therefore individually-tailored punishment according to MRPs is efficient. □

Proof of Proposition 3.6. Consider two different networks g_1 and g_2, both with peripheral players i and j of the same degree, and with subnetworks induced by all players excluding i and j the same. Suppose in g_1 i and j are connected, and in g_2 i and j are not. By construction, $\#[(N(i,rl) \cup N(j,rl)) \setminus (N(i,1) \cup N(j,1))]_{g_1} = \#[(N(i,rl) \cup N(j,rl)) \setminus (N(i,1) \cup N(j,1))]_{g_2}$. In g_2, $\#(N(i,1) \cup N(j,1)) = deg(i) + deg(j)$ whereas in g_1, $\#(N(i,1) \cup N(j,1)) < deg(i) + deg(j)$. Hence, $z^{in}_{i,j}$ for g_1 is less than $z^{in}_{i,j}$ for g_2 and so clustering makes $NWPER$ more appealing. $\square$

Chapter 4

Deceit, Group Structure, and Cooperation

Groups of people are often tasked with governing themselves. Decentralized punishment institutions maintain cooperation well when everyone knows and communicates with everyone else in a group. News spreads so rapidly that any misdeed can be punished swiftly and thoroughly. This logic is well-known and has been used to explain a wealth of examples of cooperation outside the 'shadow of the law,' from ethnic groups peacefully coexisting to traders keeping their word to ranchers minding where their livestock graze (Fearon and Laitin, 1996; Greif, 1993; Ellickson, 1991). When instead there is heterogeneity in the extent of direct communication between group members, so that some members are more peripheral than others, ensuring that everyone cooperates is more difficult. Some players are more likely to get away with offenses than others, and some players are more tempting targets than others, which ensures that universal cooperation is more difficult. In addition, some players could take advantage of others' ignorance, even if only temporary, and stand to gain from lying. This paper relates the exact network of communication in a group— a map of who communicates with whom— to how well that group can enforce cooperation *and* honesty.

Doing away with the standard assumption that everyone in a group stays perfectly informed about all others, I generalize the canonical model of inter-ethnic cooperation in Fearon and Laitin (1996) to consider the implications of an incomplete communication network (a network in which some people communicate directly with only some but not all other people). Incomplete communication networks are probably the more realistic depiction of groups with features like large populations, or time constraints, or high turnover, or limited communications technology, or sparse populations, to name a few.[1] In such cases, news may travel slowly enough that some are left out of the loop for at least some of the time. In Chapter 2 and Chapter 3, I show that when that is the case, exactly who communicates with whom affects how well decentralized punishment schemes can induce inter- and intra-group cooperation when people tell the truth.

Here I reconsider the role of networks in maintaining cooperation when people have incentives to lie as they spread information through the network. I show that opportunities to lie abound, even when networks contain nearly-but-not-quite all possible links. Groups can induce honesty and communication when they carry on as if people are truthful, eventually forgiving misdeeds, but react in outrage and terminate cooperation when they learn a lie has occurred. Some groups are better

[1] Even networks with low-cost links like Facebook friendship networks are incomplete Mislove (2009) and the network of user activity among friends is even sparser Wilson et al. (2009).

suited to enforce cooperation with such an institution. I show that groups with redundant paths between players are in a better position to verify lies early and so discourage them. Likewise, clustering among neighbors in the communication network prevents lying, making dense social cliques more valuable to honesty than a large number of neighbors.

Liars

In "environments of uncertainty" in which not everyone directly observes everyone else, communication becomes essential for enabling group enforcement of cooperation. Players can observe some actions, and can send messages to neighbors about what they have observed. If players send messages truthfully, cooperation depends on how well-integrated players are in the network Chapter 2. To consider such a case, the content of the messages is by assumption not chosen strategically– players have no option to lie about their actions or motivations. This is an admittedly strong assumption, since, as I show below, players stand to gain from lies if they could get away with them.

The assumption of truthful communication may not be so misguided. Groups may develop a norm against lying or crying wolf in other contexts and apply that norm generally even if recalculation in the new context would show gains to lying. Other contexts might contain severe consequences to lying: since links in a social network are valuable, trustworthiness is particularly important among linked individuals and deceit risks severing those valuable links (Karlan et al., 2009). An evolutionary view supports the idea that truth-telling may be adopted because of its value in at least some contexts– for example, sociobiology identifies an evolutionary mechanism favoring truthful gossip given its role in fostering reputation mechanisms for cooperation in mobile and dispersed groups (Dunbar, 1998; Enquist and Leimar, 1993). Applying norms from one context to another may be especially likely in groups in which members interact in many different ways- business interactions and every-day interactions and social interactions, say.[2] Ethnic groups have this feature, and tend to be characterized by trust among members (see, for example, Horowitz, 1985) More simply, experimental evidence suggests that when given the opportunity to gossip about others, players in cooperation games do so truthfully (Sommerfeld et al., 2007). We might also imagine that this game is embedded in a larger game left unmodeled which contains high-enough risk of being discovered and punished for lying that players play this game truthfully.

Here I consider the possibility that none of this reasoning applies, and discuss the options for sustaining cooperation *and* truth-telling in light of incentives to lie. After all, we can easily imagine groups in which players would take opportunistic behavior if they could. Groups of legislators or traders, for example, might be less cohesive and may have developed less of a norm of honesty than groups formed around social or hereditary ties.[3]

---

[2]This is the idea of multiplex networks in Sociology. See, for example, Fischer (1982); McPherson, Smith-Lovin and Cook (2001), p. 437.

[3]College students report lying in 1 out of every 3 social interactions (DePaulo et al., 1996).

Imperfect Private Monitoring with Communication

Groups of people often must make do with less-than-perfect information. In many real situations, people only interact with a subset of people, only observe a subset of people, only talk to a subset of people, or only hear about everyone else with noise. These possibilities have motivated a set of theoretical work that hunts for institutions that can induce cooperation despite missing, and sometimes erroneous information. When players do not have perfect information about everyone else, institutions incentivizing cooperation become complicated quickly, and the uncertainty renders the the usual tricks for finding cooperative equilibria useless.[4] Making players prefer cooperation to defection is difficult since the set of people who could target a perpetrator with punishment shrinks, and opportunities to profit from lying can be present.

The simplest approach in such an environment is to design an institution that does not require communication (see Kandori, 1992; Ellison, 1994). In the seminal work of Kandori (1992), it is shown that even if players are in an extremely-limited information environment in which they observe only the games in which they are involved, there exists an institution which can ensure cooperation without communication. Such an institution has players play the harshest, most alarmist strategy available, switching to defect forever once they observe a single defection. Of course such an institution is hardly robust since a single error or misinterpretation cascades into full defection and destroys cooperation forever. In many settings of interest, people *could* communicate, so the assumption of zero communication is overly restrictive for many purposes.

[4]For a careful exposition of the difficulties arising in such games, see Kandori (2002).

Acknowledging that people have the capacity to communicate, a second approach assumes players share the results of their games as widely as possible. Sometimes players are assumed to do so truthfully (as in Kandori, 1992), or are assumed to be perfectly known by some central, information-aggregating figure or mechanism (as in Anderlini and Lagunoff, 2006; Takagi, 2011). A related approach acknowledges that communication may be untruthful, but black-boxes the question of detection. In Harbord (2006), for example, players might lie but their lie will be detected (somehow) with a positive fixed probability.

Since players may face profits from lying (discussed below), for an institution to stand on its own to produce cooperation, it must also entice players to tell the truth (or account for the possibility that players may lie). When private monitoring is imperfect, a few means of inducing truth are available. One assumes that lies are either known to or detectable by some (see Calvert, 1995; Aoyagi, 2000), or that lies and the identity of the liar can be inferred Kandori and Matsushima (see 1998). In the latter approach, if the distribution of payoffs is just right, then players can use statistical inference to detect deviations, and if each player's deviation has a unique signature, players can know who the liars are. Such an approach is computationally taxing for the players. Another, computationally-easier solution uses the fact that players can determine when a lie has occurred because monitors' reports will disagree (see Ben-Porath and Kahneman, 1996, 2003). By ensuring that everyone is monitored by at least two others and punishing all who give conflicting reports, the monitors can coordinate on true reports.[5] If instead players can be insulated from the results

[5] Annen (2011) suggests that this approach may be unsatisfying since coordinating on false reports would also be an equilibrium. If any collusion is possible, we might expect this instead of the truthful

of their lies so that the veracity of their communication is independent of their own payoffs, then players no longer have a strict incentive to lie (see Compte, 1998; Lippert and Spagnolo, 2011). The interest here is in the harder case of lies from which players could strictly profit.

The approach taken here is in line with the usual solutions to private monitoring problems, though grapples with a slightly different lying problem. Here, players are observed by their neighbors and can communicate with them, who may pass along the messages to others who may pass them along, and so forth. Players have an incentive to misrepresent the history they know to defect without punishment (or at least with delayed punishment). This means their incentive to lie is greatest with second- or greater- hand information. Players' payoffs are realistically tied to their own messages (this tackles the heart of the lying problem, rather than rendering communication meaningless). I consider the possibility that players place trust in their group and presume that communication is truthful. Given truthful communication, sorting out defections is relatively simple. Any instance of untruthful communication is not tolerated, though, and any indication that lying has occurred shatters players' confidence in their group and breaks down cooperation. Such a strategy is more robust than a trigger strategy used for everything (here a capitulation, or "repentance" strategy is used for misdeeds).

This institution recognizes a qualitative difference between misdeeds and lies. Such a distinction comports with anecdotal evidence from everyday experience. A cheating significant other is usually punished more severely for lying about the scandal than

equilibrium.

committing it. Students playing a repeated prisoners dilemma can often recover from a rocky start of exchanging defections, but not from lying about what they will do. The approach here nests a game in which players assume messages are truthful in a broader game which very severely punishes lies. Lies are more difficult to detect since detection requires encountering both the true and the false information. Such an institution holds together cooperation and truthful communication, and only unravels if someone lies.

The Model

Consider two groups, A and B, each with a set of players $N = \{1, \ldots, n\}$. In each time period, all players play one round of prisoner's dilemma with a randomly assigned opponent. With probability p a player is paired with a member selected uniformly at random from the other group; with probability $1-p$ a player is paired with a member selected uniformly at random from his own group. Each pair plays a single round of the prisoner's dilemma with common payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1,1 & -\beta,\alpha \\ \alpha,-\beta & 0,0 \end{pmatrix} \end{array}$$

where $\alpha, \beta > 1$ and $\frac{\alpha-\beta}{2} < 1$. Players are rematched in each round; rounds occur indefinitely. A player's total payoff is then a stream of discounted single round payoffs. Players have common discount factor $\delta < 1$.

Let $(N, g)_A$ be a fixed simple, undirected communication network for group A with nodes N and $n \times n$ adjacency matrix g such that entry $g_{ij} = 1$ indicates the presence of a link between players i and j, $g_{ij} = 0$ indicates the absence of a link between players i and j. Let $(N, g)_B$ be the same for group B.[6] No links span $(N, g)_A$ and $(N, g)_B$.

Players can perfectly identify and recognize in-group members, but can only identify the group of out-group members. The network $(N, g)_A$ is common knowledge among players in A and $(N, g)_B$ is common knowledge among players in B; players know the shape of the other group's network but not who sits where (i.e. players only know the permutation class of the network of the other group).[7] The roster of random assignments and the actions played in each round are not observable to all players. Players observe only the matching and actions of their neighbors. This information spreads further through the network via messages passed from person to person governed by a rate of transmission, described below.

Communication

In each period t, after playing one round of prisoner's dilemma, a stage of communication occurs. Information about each game played between two players in each round is packaged in a message m containing the identity of the players (as specific as possible), the time period of the round, the actions of both players, and the motives of

[6] The assumption of identical networks for both groups is unnecessary but simplifies exposition. Both groups must be able to enforce cooperation which depends strictly on their own network. If both can do so, the equilibrium obtains.

[7] This assumption simply makes the coordination of the equilibrium more plausible. To carry out equilibrium strategies, technically players need know nothing about the other group. If they do know the shape, the know how well the other group could enforce their own, which might make it more likely that the two groups would have arrived at this equilibrium in the first place.

the players (relevant motives are determined by the strategy. For the strategy profile σ^{NWIGP} below, the relevant motive will be whether D was played out of punishment or defection.) Messages about games between same-group members i and j in t perfectly identify both players, $m_{i,j,t}$, and are sent to the neighbors of i and the neighbors of j in the communication network, e.g. to $N_g(i)$ and $N_g(j)$. Messages about games between players from different groups cannot perfectly identify the out-group opponent, and so if $i \in A$ and $j \in B$, a message $m_{i,B,t}$ is sent to neighbors of i and a message $m_{A,j,t}$ is sent to neighbors of j. A message $m_{i,j,t}$ expires after E rounds, in $t + E$. An expired message is no longer passed on. Let r govern the rate of communication spread so that player i does the following r times before $t + 1$ begins: send an unsent message about i's own game played in t and forward all unsent, unexpired messages to all of i's neighbors. This means that unexpired messages originating at i in time t are received by all players reachable in r degrees or fewer before $t + 1$.

Actions are assumed to be observable to neighbors in the communication network. That is, when player i is assigned to play j in t, i's neighbors observe that he played j and who played which action. The first round of "communication," then, can be thought of as a message in which certain contents are true deterministically- the identity of the players, the time period of the round, and the actions of both players. First, note that this still leaves something in the message to lie about- the motives of both players. Seeing that i played D and j played C is insufficient to conclude that i is now guilty. It could be that i was rightfully punishing j. More on the ability to lie below.

If all players did observe all other players (as in a complete communication network), the information about motive would be redundant in the message. All players could observe a round in which the actions are D and C and could correctly infer guilt by using their past observations to determine if the person playing C had punishment coming from a past round, which would require rounds further in the past to establish, and so on. Here, some players do not observe the full history of everyone in the game and so players rely on others' assessment of guilt in lieu of missing history. Below I discuss the difficulty of verifying guilt and the most enticing lies.

Communication spreads deterministically in that each round, everyone passes the correct number of messages to the correct recipients, and all messages intended to be sent are received unchanged. In other words, there is no misunderstanding or accidental mistakes in sending or interpreting messages. There can, however, be intentional deceit. A player can include in his message false information. The case of honest-by-assumption communication is considered in Chapter 2. Here, players may strategically choose the content of their messages.

The Value of Lying

When players choose the content of their gossip strategically, some players may have an incentive to lie. Assume that actions are observable to neighbors, so that information about who played which action is still conveyed truthfully to immediate neighbors.[8] Players would prefer to defect and not be punished, so players have an incentive to play D and claim they were punishing someone who deserved to be

[8]This assumption seems to comport with the real issue with lying. Even among a decentralized group like the Nuer in the Sudan, when players seek adjudication for disputes, disagreement is rarely over who did which action; disagreement is over whether the actions were deserved (Evans-Pritchard, 1940).

punished rather than defecting maliciously. Unless the player's neighbors know the complete history of every player in the game, some opportunity can arise in which they can not tell the difference between D the defection and D the punishment. This ambiguity is a problem because either defections are sometimes unpunished, or the group responds with a scheme which overpunishes and reduces incentives to cooperate.[9]

Specifically, a player i would like to send a message to each neighbor claiming that his current opponent j deserves punishment. Below I discuss exactly what lie would be most profitable.[10] Here I focus on the opportunity to lie. In the present setup, players observe neighbors' actions. If everyone were neighbor to everyone else (as in the complete network), players could perfectly distinguish D the defection from D the appropriate punishment because they observe all necessary information. It turns out that even small departures from the complete network admit the possibility of lying.

Call a player *discerning about* j if he can distinguish between j playing D the defection and j playing D the appropriate punishment using his observations of other players' actions.

Proposition 4.1 (Unverifiable Histories). *When a group interacts exclusively with in-group members, a player i is discerning about any $j \in N$ if and only if i has*

[9] This problem of ambiguity is well known in the study of overlapping generations models. See Takagi (2011) for a solution which makes use of a public signal that truthfully conveys information about the complete history.

[10] To preview, i's most profitable lie is one which he claims to one neighbor that he just learned a different neighbor was defected against by someone far away in the network. This must mean that the alleged culprit defected against one of i's neighbors that is rT^p degrees away rT^p periods ago. If the victim were a closer neighbor, i should have already known *and told his opponents* that j deserves punishment.

degree at least $n-2$. When a group also interacts with an out-group, a player i is discerning about any in-group player $j \in N$ if and only if player i has degree at least $n-1$.

The proof can be found in the Appendix. For the intuition, consider the condition on players in groups interacting among themselves and with an out-group. The condition says that a player must observe all other in-group players. Suppose player i observes that j play D against k who plays C. To know whether j was punishing or defecting against k, i must know whether k deserved punishment or not. This requires knowing whether k always cooperated when required and punished when required, which requires knowing whether k's past opponents always cooperated when required and punished when required, which requires knowing whether k's past opponents' past opponents always cooperated when required and punished when required, and so on recursively. Since opponents are assigned at random, in order for i to be able to distinguish good Ds from bad Ds for any possible sequence of partner assignments, i must know everyone else's history of play.

Note two features of this proposition. First, the presence of an out-group with whom players sometimes interact makes distinguishing good Ds from bad Ds more difficult. Since players observe their neighbors *and their neighbors' opponents*, so long as a player is connected to all but one other in-group members (a total of $n-2$ others), they will actually observe everyone every round. If i is connected to everyone but j and all rounds are between in-group members, j will always play someone i is connected to. In this way, i observes every in-group player's full history. If players sometimes play an out-group opponent, then part of j's history will be unverifiable to

i. That is, j will sometimes play an out-group member and i won't see what happens in this round. Players must be connected to all other in-group members (all $n-1$ others) to make motives verifiable through observation. Below I further discuss the difference between groups exclusively interacting within their own groups and those interacting with out-groups.

Second, in order to guarantee that every motive will be verifiable to every player, all players must be discerning about all other in-group players. This implies a lower bound on the number of links that must present in a network for full verifiability.

Corollary 4.1 (Verifiable Networks). *When a group interacts exclusively with in-group members, all motives are verifiable to all players only if the network contains at least $\frac{n^2-2n}{2}$ links. When a group also interacts with an out-group, all motives are verifiable to all players only if the network contains all $\frac{n(n-1)}{2}$ links.*

That is, for solo groups, players can verify everything so long as only $\frac{n}{2}$ links are missing (out of a possible maximum of $\frac{n(n-1)}{2}$) and distributed in such a way that everyone is missing only one link. For groups interacting with out-groups, players can verify everything only if the network is complete. When the networks do not contain enough links, there is always a history such that the true motive of a D played will be unverifiable to at least one player.

Lies are possible when motives cannot be perfectly verified through observation. Corollary 1 implies that very small deviations from a complete network admit the possibility of lying.[11] To cope with missing information, players can communicate to

[11] This makes the assumption of a complete network on the grounds that real world communication networks are "close enough" to complete especially suspect. Fearon and Laitin (1996) justify the assumption of perfect information within a group on the grounds that "ethnic groups are typically

fill each other in. If i doesn't observe j but k does, k could tell i what j did. If i doesn't know k but knows a mutual friend l, i could learn from l who learns from k what j did. These second- or more- hand accounts are a natural way for people to fill in information gaps.

The difficulty with second- hand information is that players have an incentive to strategically choose what their messages contain. In particular, if a player could claim to have information that someone deserves punishment so that he can play D without himself incurring punishment, he would like to do so. To induce honesty, there must be consequences to lying. In particular, it must be possible to detect lies and punishment must be strong enough to make players prefer honesty.

Detecting Lies

Since opportunities for lying are present even in very densely connected networks and since there are gains to lying, players must be able to detect lies in order to punish and disincentivize them.[12]

Before specifying a sufficient punishment strategy, consider when detecting lies is even a possibility. Messages contain a variety of content that is all chosen by the sender, including who played whom in which round, who played which action, and

marked by relatively dense social networks and low-cost access to information about other group members' behavior [...] While it is not literally true even in small ethnic groups that all members observe all interactions [...], rumor, gossip and inquiry tend to be more developed and efficacious within than between ethnic groups" (p. 721). That news can eventually reach everyone does not imply equivalence to perfect observation.

[12]Technically, players could give up on trying to detect lies and treat all messages as ignorable cheap talk. Kandori (1992) gives an example of an institution that can ensure cooperation in such a way. Given the prevalence of gossip and interpersonal communication, here the goal is to characterize an institution that preserves the usefulness of communication.

who if any are defectors. If a player faces potential punishment for a detected lie, all else equal, if there are opportunities to better conceal a lie, a player would make use of them. This intuition gives rise to the following characterization of a detection scheme:

Remark 4.1. *Detection devices which rely on aspects of the message solely manipulable by a prospective liar are suboptimal because they present a greater opportunity to conceal a lie.*

To see a straightforward example of this, consider a poorly-designed detection device which deems any message about player i to be a lie. Obviously if j wanted to construct a lie, he could do so about a player other than i to thwart detection. While this detection scheme is far-fetched and obviously suboptimal, more complicated schemes that relied on information about which players can lie are also suboptimal for the same reason.[13]

Two conditions must be met for a lie to be detectable. The first is that the liar's information must reach someone who also heard the true information. To ensure honesty in the community, this must hold for all possible liars and receivers of true information, leading to the following proposition:

Proposition 4.2 (3rd Party Reachability). *Lies are detectable on a network only if information from any pair of players can reach a third player without passing through the other of the pair. That is, for any* $i, j \in N$, $\exists k \neq i, j$ *such that there*

[13]Consider a detection device which, once a lie has been detected, claims that the identity of the liar is the first person to profit from such misinformation. Any liar could include in his lie the claim that his neighbor profited from the misinformation first and again thwart detection. More on this below.

exists a path from i to k that does not include j, and there exists a path from j to k that does not include i.

Proposition 4.2 ensures that however player i constructs a lie, the lie can eventually reach someone who will also know the true information. Even if this third party cannot tell which message is true and which is false, he will detect a conflict which indicates the presence of a lie.

Of course, the closest third party may be far away in the network, which could mean that the conflict will not be noticed for many rounds of play, the more so if the rate of information transmission, r, is small. If players only retain information for a finite number of rounds before wiping the slate clean (or plumb forgetting), the conflict may never be noticed or may be ignored. Hence, a second condition is necessary to ensure the detection of lies. Call the length of the shortest path from i and j to the nearest third party k the **Shortest Third Party Path ($STPP_{i,j}$)**, and let M be the length of memory, so that players remember information for M rounds of the game.

Proposition 4.3 (**Good Memory**)**.** *Lies are detectable on a network only if the length of memory, M, is weakly longer than the number of rounds required to reach the nearest third party of any pair. That is, lies are detectable only if*

$$M \geq \left\lceil \frac{\max_{i,j}\{STPP_{i,j}\}}{r} \right\rceil.$$

Proposition 4.3 simply states that players must remember information for a sufficiently long time so that they notice conflicting information and hence can detect lies. All lies are detectable so long as even the most difficult to monitor player's lies can be detected. Notice that the slower information spreads through the network, the greater the demands on memory. When information takes a long time to reach others (as when communications technology is poor, or news is delivered in person over long distances, etc.), the lie may not reach anyone who knows conflicting information for a very long time. The required length of memory also increases in the distance to the farthest-away-closest third party[14] by the same intuition.

Take an example which violates the condition in Proposition 4.2, shown in figure 4.1. Pairs of players on the "spokes" of this star network are reachable by a third party, like players 1 and 2 who are directly reachable by 6. Any pair of players entailing player 6, though, are not. Take the pair 1 and 6: player 1 can only reach anyone else through player 6. This violates Proposition 4.2 and gives player 6 an advantage. He could lie to someone on a spoke about someone on a different spoke without that information clashing with true information. Hence, a central figure who knows all is not sufficient to prevent lying since *he* can have incentives to lie.[15]

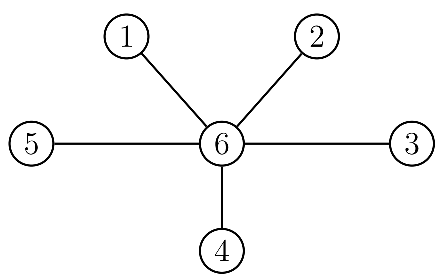

Figure 4.1: Example communication network with 6 players which violates Proposition 4.2

[14]That is, the third party closest to the pair who are farthest from their third party

[15]This observation is noteworthy in light of the special attention that "star networks" like the one in 4.1 receive in the strategic network formation literature (see Jackson, 2003). Such networks avoid redundant connections, but in settings where information needs to be verified to thwart lying, some redundancy is essential.

Likewise, consider an example which meets the condition in Proposition 4.2 but violates the condition in Proposition 4.3. Suppose the rate of communication is $r = 1$, so that news spreads a single degree each round. Suppose also that $M = 3$, so that players remember information for three rounds after the information is generated and then forget it[16], and that the communication network is as shown in Figure 4.2.

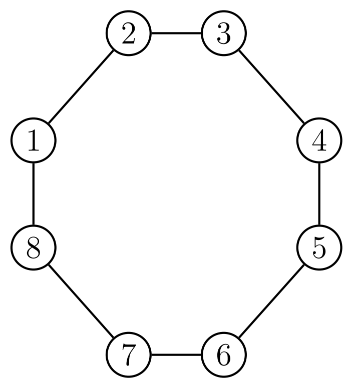

Figure 4.2: Example communication network with 8 players which meets Proposition 4.2 but violates Proposition 4.3 when $r = 1$, $M = 3$.

The network in Figure 4.2 meets the condition in Proposition 4.2 since, for any two players, a third can be reached without needing to travel through the other of the two players.[17] The nearest third party to 1 and 5 is player 3 (or player 7), which can be reached by both 1 and 5 in 2 degrees ($STPP_{1,5} = 2$). Pairs with the longest paths to third parties are neighboring pairs, like players 1 and 2. The third parties that can be reached by both 1 and 2 in the shortest distance are 5 and 6. That is, if 1 tells a lie about 2 to his neighbor 8 but 2 tells the truth about himself to his neighbor

[16] This version of memory can also be thought of as the salience of information. Notice that players do not remember information for M rounds *after they first hear the news*, but rather for M rounds *after the news is generated*. Some news is about the too-distant past to be interesting or cared about. Even with this version of memory, it is still the case that the larger is M, the more information players must keep track of, so we would expect natural constraints on M.

[17] In fact, here, *all* others can be reached without needing to travel through one of the two players.

3, those messages will conflict and the first to notice the conflict will be players 5 and 6 after 4 steps. The $\min_{i,j}\{STPP_{i,j}\} = 4$ here, and since $r = 1$, memory M must last at least 4 rounds. If news is only interesting 3 rounds after it is generated, by the time the lie reaches player 5 it is too old to pay attention to, and likewise for the truth reaching player 6. Neither would notice the conflict if news expired this quickly.

Speeding the spread of information (r), adding a link which jumps across the ring and shortens the maximum $STPP$, or increasing memory (M) would all allow the network in Figure 4.2 to meet both Propositions 4.2 and 4.3 and would make all lies detectable. These conditions do *not* guarantee that all *liars* can be identified, simply that the existence of a lie can be detected. It still may be that no player knows which of the conflicting accounts is true. As I discuss below, anonymous liars can still be incentivized to tell the truth with sufficient punishment.

Strategies to Promote Cooperation, Taking Truth for Granted

When players ignore the urge to manipulate their messages and instead communicate truthfully, the following strategy profile results in full cooperation in equilibrium under certain conditions (shown in Chapter 2):

Definition 4.1 (Network Tit-For-Tat, NWTFT). *Always punish a player (play D) known to be in punishment phase. Always cooperate with a player (play C) not known to be in punishment phase.*

Punishment phase is a status defined by the strategy profile below- in general, being in punishment phase is undesirable and the threat of entering punishment phase is what drives players to cooperate. Here, a player can "know" because of experience in his own rounds, or because of received information. A player is "not known to be in punishment phase" by a player i if i has not experienced or heard gossip that the player is in punishment phase. Now consider a strategy profile in which players play NWTFT according to the following guidelines:

Definition 4.2 (**Network In-Group Policing, σ^{NWIGP}**). *All players play NWTFT using the following definitions of status: all players begin as cooperators (not in punishment phase). A player enters (or reenters) punishment phase for T^p periods when that player (1) defects against an out-group member, (2) defects against someone not known to be in punishment phase, or (3) cooperates with someone known to be in punishment phase. A player i is known by his opponent to be in punishment phase when his opponent was the victim of (2) or (3) committed by i in at least one of the past T^p rounds or his opponent has received a message that i committed (1), (2) or (3) in at least one of the past T^p rounds.*

Punishment here takes the form of capitulation (playing C against a punisher playing D) as it does in Fearon and Laitin (1996) and Calvert (1995). That defectors may be required to atone for their offense by capitulating or repenting for a certain number of rounds has the nice property that players like to punish defectors, and squares with real world observed punishment regimes (Harbord, 2006). The above is really a set of strategy profiles, each with a fixed length of punishment phase T^p. Chapter 2 shows that some sparse networks and slowly communicating groups can only maintain full cooperation with a sufficiently high T^p.

Full cooperation is possible, assuming truthful communication, when the following conditions are met:

Proposition 4.4 (**Full Cooperation ASSUMING Truth-Telling**). σ^{NWIGP} *is sequentially rational for game G with networks* $g_A = g_B$ *iff*

$$\delta^{T^p} \geq \max \left\{ \frac{\alpha - 1}{(1-p) z^{out}_{min,rT^p}(\beta + 1)}, \frac{\beta}{(1-p) z^{in}_{min,rT^p}(\beta + 1)} \right\}$$

and

$$p < \min \left\{ \frac{z^{in}_{min,rT^p}(1+\beta) - \beta}{z^{in}_{min,rT^p}(1+\beta)}, \frac{z^{out}_{min,rT^p}(1+\beta) - \alpha + 1}{z^{out}_{min,rT^p}(1+\beta)} \right\}$$

where z^{out}_{min,rT^p} and z^{in}_{min,rT^p} are the probability that the least-punishable defection will be punished by an in-group member at the end of the punishment phase (in $t+T^p$) for defecting in t against an out-group member and in-group member, respectively.[18]

The binding player here is the one least observable by other players; in other words, the most "peripheral" player. In figure 4.3, player 4 is most peripheral. If news travels slowly enough and the punishment phase is short enough, he can expect to defect and face little punishment *even if news about him travels truthfully.*

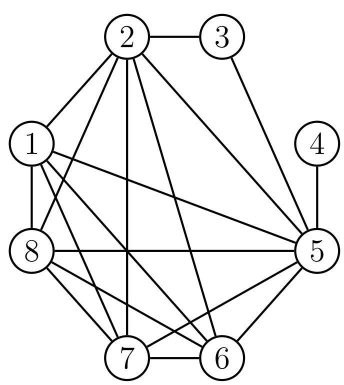

Figure 4.3: Example network where player 4 is most peripheral and so is the binding player for full cooperation, assuming truthful communication.

[18]The proof of Proposition 4.4, a discussion of consistent beliefs to extend the strategies to a sequential equilibrium, and the full specification of the probabilities z^{in}, z_{out} as functions of network characteristics can be found in Chapter 2.

Strategies to Promote Cooperation AND Honesty

Players can induce cooperation and honesty by punishing lies as well as defections. The complication is that when a network is missing even a few links (as per Corollary 1), sorting out defections from punishment is difficult when players can strategically choose their messages. Even when the conditions in Propositions 4.2 and 4.3 hold, players may have difficulty determining *who* the liar is. Standard solutions like shoot-the-messenger approaches (see, for example, Ben-Porath and Kahneman (1996)) become quickly intractable when messages are passed along second- and greater-hand since there can be a string of messengers passing news before it reaches conflicting news.[19] The approach taken here blends the grim approach of Kandori (1992) with shoot-the-messenger approaches to capture an institution seemingly undertaken by real-world groups.

In particular, players carry on assuming that other players are trustworthy and react strongly if they receive evidence that their trust was taken advantage of. Players effectively play a game in which they naively trust each other which is embedded in a broader game in which violations of this trust trigger the end of cooperation. Players are willing to trust until they learn a liar is afoot.

Consider the following strategy profile. It is a modification of σ^{NWTFT} above in which players also choose the content of their messages and are sensitive to conflicting information.

[19]When messages are passed along indirectly, it becomes difficult (and sometimes impossible) for players to determine where the actual punishment for lying started, how long to carry out punishments, and how to coordinate these decentralized efforts so that they eventually end and players can return to cooperating.

Definition 4.3 (Network In-Group Policing with Messages, σ^{NWIGPM}). *All players begin in "trustworthy" phase and regard all messages as truthful. So long as no information conflicts, play NWTFT using the following definitions of status: all players begin as cooperators (not in punishment phase). A player enters (or reenters) punishment phase for T^p periods when that player (1) defects against an out-group member, (2) defects against someone not known to be in punishment phase, or (3) cooperates with someone known to be in punishment phase. A player i is known by his opponent to be in punishment phase when his opponent was the victim of (2) or (3) committed by i in at least one of the past T^p rounds or his opponent has received a message that i committed (1), (2) or (3) in at least one of the past T^p rounds. At the end of a round, send all neighbors a true message containing the time, opponents, actions, and motives of the most recent round, and pass truthfully all unexpired, received messages r times. If a player receives information that conflicts with any of his previously known information or observes a "jaded neighbor," he switches to "jaded" phase and plays ALLD.*

In a nutshell, players presume truthful messages and punish defections in the usual way. Once someone learns that all have not been truthful, jadedness sweeps through the network, terminating cooperation.

Note that such an alarmist strategy does not require knowing the identity of the liar. A player need only know that a lie has occurred, not the identity of the liar, to trigger the grim punishment. Being a grim punishment, a cooperative equilibrium will not be fully robust to mistakes. If someone mistakenly lies or sounds the liar alarm by switching to jaded phase, cooperation is terminated. However, such an equilibrium *will* be robust to mistakes in behavior. If someone accidentally plays D when they were supposed to play C, players observe misbehavior that deserves to be punished. So long as news spreads consistently about the defection, the liar alarm does not sound and cooperation can resume after the erring player incurs a finite punishment phase. Such an equilibria is more robust than that induced by fully-grim punishment as in Kandori (1992).

Observation suggests that people behave similarly to the prescription in σ^{NWIGPM} in environments in which truth is difficult to verify. Take romantic relationships. Not every late evening at the office can be verified, nor is a partner necessarily incentivized to be faithful at all times. Anecdotally, there does seem to be a difference between reactions to misbehavior and reactions to *lying* about misbehavior. The latter tends to elicit a stronger reaction, as in "worst of all, you weren't honest with me."

Likewise, any in-class run of a finitely repeated dilemma results in some students colluding despite instructions to the contrary. Pairs which never collude but get off to a rocky start can return to cooperating. Not so for the pairs who try to collude but take advantage of the other's trust. Once the communication is revealed to be a lie, game over- mutual defections finish the game. Comparing reactions to misbehavior to reactions to lies is ripe for an experimental setup. More on this below.

Most Profitable Lies

Since players pass along information second-hand, a player considering lying has options. He could lie about himself, or he could lie about other people. A liar benefits from playing D against a C without incurring punihsment for his misbehavior, but faces costs from tripping the alarm and converting his group to disillusioned non-cooperators.

A prospective liar's costs are greater when he trips the alarm more quickly, and when the cascade of jaded players unfolds more quickly. To delay tripping the alarm, the liar must maximize the time until his false information collides with the true information. Recall that player i is only supposed to punish people he *knows* deserve punishment, and also that player i passes along information he knows to his neighbors. If player i is assigned to play j, defects against him and tries to tell his neighbors that j had it coming, his neighbors should know i knew this if i supposedly learned about j's treachery long enough ago. The only way i's neighbors would think i might know about j's treachery without themselves knowing i knew it is if i just learned it and hasn't had a chance to pass the information along.

Remark 4.2. *A lie which only misrepresents one's own motive is not as profitable as a lie which misrepresents one's own motive and lies about someone else's history of play.*

In other words, the best lie i could make is to claim that his newly assigned opponent j defected against someone in the past such that i just learned about the defection at the end of last round and didn't yet share it with his neighbors. Any other lie would either create conflict immediately or earn i punishment for punishing someone he did not know to deserve punishment.

Lemma 4.1 (Most Profitable Lie). *Player i's most profitable lie is of the form: i tells neighbor j that i's current opponent defected against someone tk degrees away* *in round $t-k$ so long as j does not sit on the path between i and the supposed victim tk degrees away.*

If player i can tell his most profitable lie to a neighbor who does not sit on the path between i and the supposed victim tk degrees away, he can avoid punishment from the lied-to neighbor and the neighbor's neighbors and so on. If he can tell such a lie to all neighbors, which might entail telling different neighbors different lies, he can avoid any immediate punishment from players other than the victim (who experiences a defection against him and will start punishing the misbehavior).[20]

Proposition 4.5 (Most Profitable Bundle of Lies). *A player's most enticing set of messages about a particular opponent is the set which tells a most profitable lie to each neighbor.*

So the most likely liar is the player i who has an opponent that could be assigned such that i can tell a most profitable lie about the opponent's history and i's message will take as long as possible to collide with information from either the opponent or the opponent's alleged past victims.

Remark 4.3. *There exists an opponent which, if assigned to i, weakly maximizes i's gain to his most profitable bundle of lies and so is i's most tempting opponent to lie about. The person who stands to gain the most from his most profitable bundle in his most tempting pairing is the binding player: group honesty hinges on keeping this player telling the truth.*

[20]The victim experiences a misdeed and not a lie- he will punish the defection with finite punishment, not set off the liar alarm.

Take a simple example, using the stylized network in Figure 4.4 and suppose that $r = 1$ so that news travels one degree after each round. Consider what happens when a player lies– he defects against someone who then wants to punish his bad behavior, and he lies to all those watching, which, if the conditions in Propositions 4.2 and 4.3 are met, will eventually reach someone who knows conflicting information and trigger the liar alarm. In this example Player 1 passes along information to 9 and 2, his two neighbors. The only way 1 would know that his current opponent deserves punishment without already having told 9 and 2 is if 1 *just* learned it, which would mean here that the offense targeted his other neighbor. 1 can tell 9 that he just learned 2 was victimized by the current opponent, and can tell 2 that he just learned 9 was victimized by the current opponent.

The most profitable opponent to lie about in this example is a neighbor and that neighbor's neighbor. Doing so maximizes the time to information conflict without increasing the expected punishment for his misdeed. For example, when assigned to play 2 in t, 1 can lie to 9 that 2 defected against 3 in round $t-1$, and tell the truth to 2, that 1 is defecting against 2. This lie will be discovered when the true information about 2 and 3 in $t-1$ reaches someone who receives the lie 1 started about 2 and 3 in $t-1$. Here, the conflict occurs 3 rounds after the lie. Any other lie takes a weakly shorter time to collide.

Because true information creates a risk of colliding with false information and revealing a lie, situations with fewer true messages weakly reduce this risk. Messages spread deterministically between in-group players, but not between the in-group and the out-group. This leads to the following consequence of having inter-group relations:

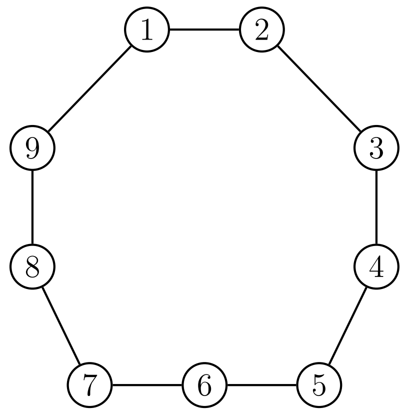

Figure 4.4: Any player's most profitable in-group lie is to lie about a neighbor to the other neighbor. 1 can tell 9 that 2 defected against 3 in $t-1$ when $r=1$ so that 1 can defect against 2.

Lemma 4.2 (Inter-group Relations Increase Opportunities to Lie). *The presence of an out-group with whom people sometimes interact but from whom no information reaches the in-group weakly increases the benefeis to lying.*

Corollary 4.2 obtains because when all interactions occur within a group, a lie about a past round must be about two people. These two have spread information about their history that will conflict with the false history the liar spreads. When some interactions occur between groups, a lie need only entail one in-group member and hence one person who spreads information which will conflict with the lie.

In Figure 4.4, 1 can tell 9 that 2 defected against an out-group member (instead of against 3 as in the earlier example). Now 2 spread the truth, 9 accidentally spreads the lie, and the information will conflict in 4 rounds (as opposed to 3).

Similar logic suggests that players whose neighbors are connected to each other (clustered) have a more difficult time lying than players whose neighbors are not connected to each other.

Lemma 4.3 (Clustered Neighbors Thwart Lying). *A player whose neighborhood is not a clique*[21] *faces a weakly more profitable lie than a player whose neighborhood is a clique.*

The best lie tells neighbors a player has private information that another player deserves punishment. If all neighbors are connected, though, they receive all information at the same time as the rest of their neighbors. Player i has no sources that the other neighbors do not also independently have. i can't claim to have privileged information that his opponent should be punished without his lie being detected immediately. In fact, this reasoning suggests that clustering matters more than degree.[22]

Corollary 4.2 (Clustering Matters More than Degree). *The presence of a small number of clustered neighbors does more to decrease the gains from lying than the presence of a large number of disjoint neighbors.*

Take Figure 4.5 in which player 1's neighbors are now also neighbors to each other:

If in t player 1 tries to tell player 9 that 2 defected against the outgroup in $t-1$, this information will immediately clash with the true information 2 provided at the end of $t-1$. Any other player whose full neighborhood is not a clique can provide a lie that is not immediately detectable. Even a player with some neighbors forming a clique, like player 2 with neighbors 1 and 9 connected, can avoid detection for a while by telling the connected neighbors one lie and the other neighbors (3) something

[21] All players in a clique are connected to all other players. Here, a player's neighbors are all connected to each other.

[22] It can even be shown that adding a disjoint neighbor to a player with otherwise completely clustered neighbors *increases* his incentive to lie.

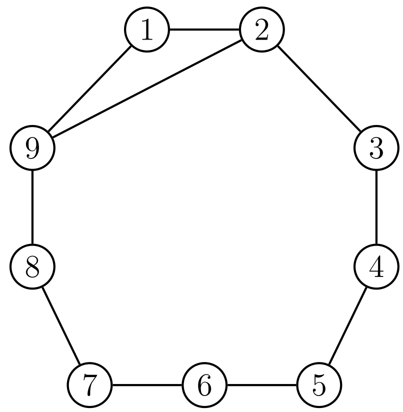

Figure 4.5: Player 1's neighborhood is a clique, so 1 will be immediately caught lying if he tries to tell any of his neighbors he has information that they have not received.

else.[23]

The usefulness of clustered neighbors is similar to the result in Ben-Porath and Kahneman (1996) which finds that the minimum number of monitors observing each player required to obtain truth-telling in equilibrium is 2. Here, having 2 monitors may be insufficient- players are all observed by 2 others in Figure 4.4, and yet there are gains to lying that could outweigh the costs (shown in the next section). The problem is not with keeping the monitors in line, but with verifying the information provided to the monitors. Clustering makes this possible.

Truth and Cooperation in Equilibrium

Under certain conditions, the prospective liar who would profit the most from his most profitable lie can be enticed to tell the truth (and with him, all other prospective liars). If all players also prefer to cooperate assuming truthful messages, and lies are detectable, then the strategy profile σ^{NWFTFM} results in true messages and fully

[23]Such a lie would do best if it told neighbors 1 and 9 that the opponent defected against the other neighbor, 3, and told neighbor 3 that the opponent defected against one of 1 or 9.

cooperative behavior in equilibrium.

Proposition 4.6 (Fully Honest and Cooperative Equilibrium). *σ^{NWTFTM} is sequentially rational for game G given networks $g_A = g_B$ if all of the following hold:*

For any $i, j \in N, \exists k \neq i, j$ such that there exists a path from i to k that does not include j, and there exists a path from j to k that does not include i (4.1)

$$M \geq T^p + \left\lceil \frac{\max_{i,j}\{STPP_{i,j}\}}{r} \right\rceil \tag{4.2}$$

$$\delta^{T^p} \geq \max\left\{ \frac{\alpha - 1}{(1-p)z^{out}_{min,rT^p}(\beta+1)}, \frac{\beta}{(1-p)z^{in}_{min,rT^p}(\beta+1)} \right\} \tag{4.3}$$

$$p < \min\left\{ \frac{z^{in}_{min,rT^p}(1+\beta) - \beta}{z^{in}_{min,rT^p}(1+\beta)}, \frac{z^{out}_{min,rT^p}(1+\beta) - \alpha + 1}{z^{out}_{min,rT^p}(1+\beta)} \right\} \tag{4.4}$$

$$\alpha - 1 \leq y_{cont,i}\big|^{\infty}_{t+T^E_i+1} - y_{pre,i}\big|^{T^X_i}_{t} - y_{post,i}\big|^{t+T^E_i}_{t+T^X_i+1} \;\forall i \in N \tag{4.5}$$

(1) and (2) are from Propositions 4.2 and 4.3, and (3) and (4) are from Proposition 4.4. T^X_i is the number of rounds before i's most profitable lie reaches someone who knows conflicting information and T^E_i is the number of rounds before i's most profitable lie results in full jadedness (the collapse of cooperation). Here, $y_{cont,i}|^{\infty}_{t+T^X_i-1}$ is the value i receives from continuing the game indefinitely starting from the round after a conflicting message would be discovered. $y_{pre,i}|^{T^X_i}_{t}$ is the value to i of taking full advantage of the peiod between the lie and its detection, and $y_{post,i}|^{t+T^E_i}_{t+T^X_i+1}$ is the value to i of taking full advantage of the period after the lie's detection but before complete breakdown of cooperation. Once again, the assumption of same networks

in the two groups is unnecessary; the above conditions simply must hold within both groups, whatever their networks look like. A more precise characterization of these terms can be found in the Appendix.[24]

In words, Proposition 4.6 says that full cooperation and honesty are sequentially rational if (1) all pairs can reach an independent third party, (2) messages remain salient long enough, (3) & (4) conditions are right to induce cooperation if messages were truthful, and (5) the gains to continuing the game outweigh the gains from lying. The proof of Proposition 4.6 can be found in the Appendix, as well as a discussion of consistent beliefs which extend the behavior to a sequential equilibrium.

Condition (5) keeps potential liars telling the truth. Clearly whether (5) is satisfied or not depends on properties of the network and players communicating on it.

Corollary 4.3 (**Keeping Liars Honest**). *Condition (5) is easier to satisfy as detection of lies happens more quickly (closer third-party observers to shrink T^X), collapse occurs more quickly (the network has smaller diameter to shrink T^E)[25], and players value future gains more highly (higher δ).*

Consider a simplified numerical example to confirm the existence of such an equilibrium for the group as pictured in Figure 4.6. It can be shown that, if members of the group occasionally interact with an out-group not pictured, the above conditions are met when $T^p = 2$, $r = 1$, $M = 6$, $\delta = .95$, $\alpha = 1.01$ and $\beta = 1.015$. Once again the most profitable lie is one in which a player lies to one neighbor that the

[24] Proposition 4.6 assumes that there are both intra- and inter-group interactions. A group playing σ^{NWTFTM} when there are no inter-group interactions would be fully cooperative in equilibrium under the same conditions, setting p to 0 in condition (3) and ignoring condition (4).

[25] Collapse occurs in $\lceil \frac{Diam}{r} \rceil$, so that if a lie occurs in t, cooperation will be expunged by $t + \lceil \frac{Diam}{r} \rceil$.

other neighbor defected against the out-group. That makes the time to information conflict, $T^X = 4$ and the time to the end of cooperation $T^E = 7$. Since $T^X > T^p$, a player could expect the group to be cooperative if he did not lie, earning $\sum_5^\infty \delta^{T^p}$. The amount a liar could expect to gain in the time before misinformation is detected is bounded above by earning α each round, and likewise for the amount gained in the time until cooperation ends.[26] The amount foregone by triggering the liar alarm outweighs even these optimistic gains. From Chapter 2 we can calculate that the probability of punishment at the end of punishment phase when defecting against an outsider, z^{out}_{min,T^p} here is equal to $\frac{1}{2}$ and the same when defecting against an in-group member, z^{in}_{min,T^p} is $\frac{5}{8}$. If players encounter the out-group infrequently enough, say with $p = .1$, players prefer to cooperate with everyone when messages are truthful. Since memory is long enough and any pair can be reachable by a third independent player, $\sigma NWTFTM$ results in a fully cooperative and honest equilibrium.

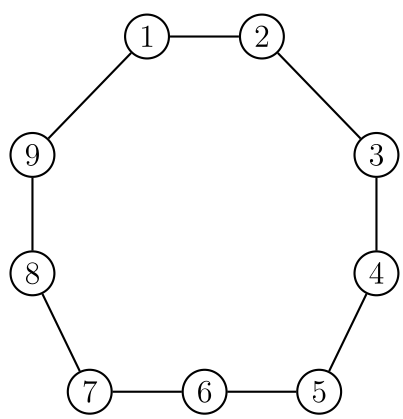

Figure 4.6: A fully cooperative and honest equilibrium is possible even for the above network.

Consider what could happen if condition (1) were violated. Figure 4.7 shows one

[26]Of course, the player would be overly optimistic if he expected this full payoff- there is a chance he would encounter the player he wronged and fail to gain α from him, for example.

such network, reproduced from above.

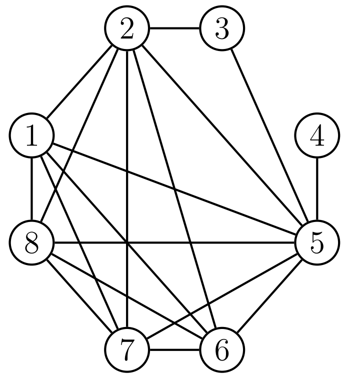

Figure 4.7: Example network where information from player 4 only reaches others via player 5.

Player 5 controls access to information about player 4. Without a third party reachable independent of player 5, 5 can claim that 4 defected against the out-group, say, and no news to the contrary ever reaches anyone else. 4 could punish 5 in return, but 4 could always spin this to his neighbors as a defection against him and avoid further punishment. Controlling access points to information opens opportunities to lie.[27]

Discussion

The above shows that groups for which verifying information through observation is difficult can still maintain cooperation and honesty with decentralized institutions. Players who stand to gain from lying are discouraged from doing so when the costs of lying will be realized and quickly enough.

One interesting result is really the absence of a result: a player's degree is not

[27]The same is true for "brokers" or "bridges" in a network, the players who connect two otherwise disjoint components. These players control which information reaches which component of the network and have a greater opportunity to deceive.

what matters. That is, it is not the number of connections a player has that makes the difference per se but how those connections are arranged. Increasing the number of neighbors for a player has a very small effect on the incentive to lie, and the net effect can even be to increase the value of lying.[28] Typical monitoring stories focus on the strength of oversight, but here it is not the volume of oversight that matters. Instead, here what matters is the extent to which the overseers receive information from the same sources that the person being monitored does and at the same time.

When a player's neighbors are all themselves connected, lying is detected immediately. This is because lies take the form "in the past, person j defected against person k." If neighbors all receive the same accounts of the past that i receives and at the same time, they will know immediately if i tries to change his telling of the past for his own gain. A small number of clustered neighbors goes farther toward preventing lying than a large number of un-clustered neighbors.

The above results suggest that there are some social structures which are more conducive to honesty and cooperation than others in environments of uncertainty. Given evolutionary pressure to cooperation, we might expect an evolutionary preference for certain structures or features of groups over others.

For example, as per condition (1) in Proposition 4.6, groups without ubiquitous third-party reachability (i.e. groups in which one person can control access to others' information) pose problems for honesty and therefore cooperation in equilibrium. As seen in Figure 4.7 person tenuously connected to a group relies on the good nature

[28]Increasing the degree of a player without increasing the total number of links must mean that paths to a third party weakly decrease in length which weakly increases the cost to lying. Increasing the degree also increases the probability that a player will be assigned to play one of his most tempting victims to lie about.

of their sole contact. Likewise, someone serving as a bridge between two sections of a network that would be disjointed without the bridge's ties control information and make lying to one section about people in the other section a tempting possibility. Given the difficulty of enforcing honesty in such settings, we might expect few examples of true bridges.[29]

Similarly, we should expect greater pressure to build clusters of friend groups rather than amass a large number of possibly unacquainted friends. Clustered friends can verify information sources and encourage truth-telling in ways that a laundry list of isolated friends cannot. It may be no surprise that human networks (as opposed to networks of objects or ideas) tend to feature high clustering (Watts and Strogatz, 1998).

If groups rely on in-group policing mechanisms, we might also expect pressure to interact almost exclusively with in-group members. The presence of an out-group provides opportunities to lie by siphoning off potential sources of information. While both parties to an interaction share information about it when meeting in-group members, an out-group member's account of the interaction never permeates the in-group. This does not create opportunities to defect against the out-group – since everyone is always instructed to cooperate in those interactions and actions are observable, no lie could excuse an observed D against the out-group– but does create opportunities to claim that *others* defected against the out-group.

Whether groups truly make use of the institution described here is an open ques-

[29]Granovetter (1973) makes a similar claim, observing that when bridges do exist, they should not be comprised of a strong tie. The reasoning here is different, and suggests another reason why groups may experience pressure to build in redundancy and to close triads.

tion. A first step toward empirically verifying the institution could be taken in the laboratory. A key component to the institution is a difference in reaction to misdeeds and to lies. Players are willing to eventually excuse misdeeds but not lies. Anecdotal evidence suggests that people do tend to react more strongly to lies, but more rigorously verifying this reaction is possible and poses a promising direction for future research.

Conclusion

When communities are tasked with enforcing behavior, information is essential. People must know who misbehaves and who doesn't. When doling out punishment looks like misbehavior, and community members are supposed to punish misdeeds committed against anyone, the informational demands are extremely strong. If only a few links are missing from the network, rightful punishment cannot always be distinguished from wrongful actions.

This suggests that lying is an important hurdle to cooperation in environments lacking perfect information. This also suggests that assuming complete networks on the grounds that real networks are "close enough" may in fact be a big leap that overlooks difficulties arising in networks very close to complete ones.

I have established when lies can be detected, characterized the most profitable lies that must be disincentivized if honesty is to obtain, and established an institution that can sustain both honesty and cooperation. The institution embeds trusting behavior in a game which treats lies gravely. Players can eventually forgive misbehavior but not lies. This institution is more robust than institutions which respond gravely to both actions and lies.

Here, players may lie because they stand to gain personally. Another lying scenario is possible. Someone may lie because they are interested in changing the group behavior. For example, someone may lie because they have a vendetta against someone or want to incite conflict with another group. While this case isn't taken up here, an interesting area of future research entails the role of networks at thwarting or magnifying incendiary rumors.

Appendix: Proofs

Proof of Proposition 4.1. If i observes j playing D against player k in t, i must determine whether k deserved punishment, which depends on k's actions in the last T^p rounds. Any single round can establish k's guilt, so if i observes all but 1 of k's last T^p rounds, $\exists$ history such that k is guilty because of the unobserved round and appears innocent otherwise. Therefore the maximum number of people who must be observed to know k's relevant actions is T^p. k's guilt against any of the T^p opponents depends on *their* guilt and hence their past T^p opponents, any one of whom could determine their guilt. Their status depends on *their* past T^p opponents, and so on back. In other words, the maximum number of people who must be known is $(T^p)^t$. Since the game is infinitely repeated, $t \rightarrow \infty$, there is a point in the game beyond which the maximum number of people who must be known exceeds $n-1$ and $\exists$ history of sufficiently-mixed round matches which maximizes the required number of people known. Therefore, $n-1$ people must be known to safeguard against a history in which guilt cannot be determined. When matches are all in-group, $n-2$ links suffices since the one person not linked to will always be matched to someone connected to i. When some matches are with the out-group, all $n-1$ links are required since any person not connected could be matched with the out-group and escape observation. $\square$

Proof of Corollary 4.1. The proof follows immediately from Proposition 4.1 and properties of simple, undirected networks. For all i to be discerning about all possible opponents in exclusively in-group interactions, all must be connected to $n-2$ others. The largest complement network in which all are connected to at most 1 other has $\frac{n}{2}$ links, so the smallest network which excludes the complement network contains $\frac{n(n-1)}{2} - \frac{n}{2}$ links. When some interactions are with an out-group, all $\frac{n(n-2)}{2}$ links must be present. $\square$

Proof of Proposition 4.2. Players can detect a lie by knowing the truth through observation and hearing a lie, or by not knowing the truth but perceiving conflicting information. By Proposition 4.1, all can verify truth if they have degree $\geq n-2$, which implies all pairs have paths to an independent third party, establishing necessity for the first detection method. If truth cannot be verified through observation, a lie can be detected if some player receives both the lie and the conflicting true information. Suppose $\exists$ pair i, j such that all paths to all other players from j go through i. Then i can prevent lies about j from reaching conflicting information. If, however, $\exists$ a player l who can be reached by j without crossing i and from i without j, neither player can prevent conflict. $\square$

Proof of Proposition 4.3. The nearest 3rd party is the first location of information conflict, i.e. the first opportunity to discover the presence of a lie. If players forget information before it even reaches this 3rd party, conflict will never be noticed and the lie will be undetected. □

Proof of Lemma 4.1. Taking advantage of a lie in the future rather than immediately decreases the window of exploitation and makes the gain probabilistic (since a player may or may not be randomly assigned to play the person lied about), so a liar cannot do better than to lie about a current opponent.

Given the same punishment for deviating, no deviation is as profitable as playing D when instructed to play C, so a liar cannot do better than to claim his opponent deserved punishment.

Lying about a round involving both players more than rk degrees away in $t - k$ means lying about players whose information should not have reached i yet, so the lie is immediately detectable to i's neighbors. Lying about a round involving a player fewer than rk degrees away in $t - k$ means lying about a round i already told his neighbors about, so again the lie is immediately detectable. Hence, the best lie is about a round that took place in $t - k$ which involved one player rk degrees away from i.

If neighbor j is on a path $\leq rk$ degrees from the alleged victim, j already knows the truth when i tells him the lie, so the conflict is detected immediately.

□

Proof of Proposition 4.5. The proof is immediate from Lemma 4.1. Each player does best by picking a most profitable lie to tell each neighbor. □

Proof of Lemma 4.2. Take i's most profitable lie when interactions are in-group only, which entails a person rk degrees away from i and another ingroup member in an alleged round in $t-k$. The lie can be modified to entail the person rk degrees away and a member of the out-group when an out-group exists. Such a lie is no more detectable but entails only one source of the true information (as opposed to two sources when both subjects of the lie are in-group members). Decreasing the number of sources of information that can conflict with the lie never increases the costs of a lie, all else equal. □

Proof of Lemma 4.3.] Suppose i's neighbors, $N(i)$, are all connected to each other. This implies that any path from a player rk away from i can reach all $n(i)$ at least as quickly as i. Therefore, any lie which claims i knows something $N(i)$ do not is discoverable immediately. If $\exists$ a neighbor $l \in N(i)$ who is not connected to the other neighbors $N(i) \setminus l$ and $N(i) \setminus l$ form a clique, then $\exists$ lie i could tell which is not discoverable immediately. By construction, information from the clique and the other neighbor will not collide until, at the earliest, one additional round. □

Proof of Proposition 4.6. (1) Given that players spread messages to neighbors and detect lies based on conflicting information, the sufficiency of (1) is clear.

(2) From Proposition 4.5, we know the most profitable lie entails a lie which delays conflict. Players can take advantage of short memory and construct a lie about someone rT^p away, who supposedly defected in $t-T^p$. Such a lie ensures that people forget it after i says it while still allowing him to punish within the punishment window. Players remember long enough to detect any profitable lie when (2) holds.

(3) & (4) When players are in trusting phase, these conditions are sufficient to discourage any deviation in behavior. The proof can be found in Chapter 2.

(5) Players prefer to keep cooperating and avoid triggering jadedness when the

gains to the most profitable lies are smaller than the costs to remaining trustworthy. If i complies and refrains from lying, he expects to earn

$$1 + \sum_{l=1}^{\infty} \delta^l \left[(1-p)\left[prob_l(c^*) + prob_l(d^*)\alpha\right] + p\right]$$

where $prob_l(c^*)$ and $prob_l(d^*)$ are the believed probabilities of being randomly matched with a cooperator and a defector, respectively, in $t+l$.

If instead i lies and takes full advantage of the opportunities presented by lying and the impending end of the game, he expects to earn

$$\alpha + (y_{exploit} - y_{pun})|_{t+1}^{t+T^X} + (y_{exploit} - y_{pun})|_{t+T^X+1}^{T^E} - \sum_{l=T^E+1}^{\infty} \delta^l \left[(1-p)\left[prob_l(c^*) + prob_l(d^*)\alpha\right] + p\right]$$

where, once i breaches trust, he faces possible gains from exploiting the impending end of the game and costs from punishment he expects to incur from his victim who thinks he committed a misdeed and spreads the word that i deserves finite punishment. The exact specification of these values is not essential for present purposes- for a numerical example confirming existence, see section 5.

$(y_{exploit} - y_{pun})|_{t+1}^{t+T^X}$ are the net gains from exploitation before anyone detects the lie and $(y_{exploit} - y_{pun})|_{t+T^X+1}^{T^E}$ are the same after the lie is detected but before jadedness fully overtakes the group. These terms can be rewritten $y_{pre}|_{t+1}^{t+T^X}$ and $y_{post}|_{t+T^X+1}^{T^E}$ for convenience. Likewise, the value to continuing the trustworthy cooperation game, can be rewritten $y_{cont}|_{t+T^E+1}^{\infty}$. A player prefers to play honestly and keep players in trusting phase when (5) holds.

Hence, conditions (1) through (5) establish sufficiency for sequential rationality.

The behavior in equilibrium is of greater interest than the beliefs in equilibrium. The above shows that σ^{NWIGPM} forms half of a sequential equilibrium since the strategy is sequentially rational in any information set with any beliefs over nodes in the information set. Consider a set of beliefs, μ, which places probability one on all events expected in equilibrium, probability zero on out-of-equilibrium events, and which are updated according to Bayes' rule. Now consider a perturbation of these beliefs in which players are assumed to tremble with small independent probability $\epsilon > 0$. Given the perturbation, all information sets are reached with positive probability and Bayes' rule pins down beliefs everywhere. Let μ^* be the limiting beliefs derived from Bayes' rule as $\epsilon \to 0$. Since sequential rationality holds for any specification of beliefs given the above conditions, it holds for μ^*. Hence, the assessment (σ^{NWIGPM}, μ^*) is a sequential equilibrium.

$\square$

Proof of Corollary 4.3. $y_{pre}|_{t+1}^{t+T^X} \geq 0$ and $y_{post}|_{t+T^X+1}^{T^E} \geq 0$ since player i could always block defections against himself while the game is ending. Therefore, (5) is weakly easier to satisfy as T^X decreases because smaller T^X means fewer opportunities to gain from exploitation. The same applies to T^E. As the future is valued more highly, the continuation value increases faster than the two short term gains from the impending end of cooperation, making (5) easier to satisfy as δ increases. $\square$

Bibliography

Abreu, D., D. Pearce and E. Stacchetti. 1990. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica: Journal of the Econometric Society* pp. 1041–1063.

Anderlini, L. and R. Lagunoff. 2006. "Communication in Dynastic Repeated Games:Whitewashes and Coverups." *Rationality and Equilibrium* pp. 21–55.

Anderson, T.L. and P.J. Hill. 2004. *The Not So Wild, Wild West: Property Rights on the Frontier*. Stanford University Press.

Annen, K. 2011. "Lies and Slander: Truth-Telling in Repeated Matching Games with Private Monitoring." *Social Choice and Welfare* 37(2):269–285.

Aoyagi, M. 2000. "Communication Equilibria in Repeated Games with Imperfect Private Monitoring.".

Aumann, RJ and L. Shapley. 1976. "Long-Term Competition: A Game Theoretic Analysis." *mimeo* .

Bandiera, O. 2003. "Land Reform, the Market for Protection, and the Origins of the Sicilian Mafia: Theory and Evidence." *Journal of Law, Economics, and Organization* 19(1):218.

Banerjee, A., A. Chandrasekhar, E. Duflo and M. Jackson. 2011. "Social Networks and Microfinance." *Unpublished Manuscript, MIT* .

Barr, A., J. Ensminger and J.C. Johnson. 2010. Social Networks and Trust in Cross-Cultural Economic Experiments. In *Whom Can We Trust?* Russell Sage Foundation Press.

Bellair, P.E. 1997. "Social Interaction and Community Crime: Examining the Importance of Neighbor Networks." *Criminology* 35:677.

Ben-Porath, E. and M. Kahneman. 1996. "Communication in Repeated Games with Private Monitoring." *Journal of Economic Theory* 70(2):281–297.

Ben-Porath, E. and M. Kahneman. 2003. "Communication in Repeated Games with Costly Monitoring." *Games and Economic Behavior* 44(2):227–250.

Blondel, V., G. Krings and I. Thomas. 2010. "Regions and Borders of Mobile Telephony in Belgium and in the Brussels Metropolitan Zone." *Brussels Studies* 42(4).

Bo, P.D. 2005. "Cooperation Under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games." *The American Economic Review* 95(5):1591–1604.

Boehm, C. 1984. *Blood Revenge: The Anthropology of Feuding in Montenegro and Other Tribal Societies*. University Press of Kansas Lawrence.

Borgatti, S.P., C. Jones and M.G. Everett. 1998. "Network Measures of Social Capital." *Connections* 21(2):27–36.

Bott, E. 1971. "Family and Social Network: Roles, Norms, and External Relationships in Ordinary Urban Families..".

Burt, R.S. 2005. *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press, USA.

Calvert, R.L. 1995. Rational Actors, Equilibrium, and Social Institutions. In *Explaining Social Institutions*, ed. J. Knight and I. Sened. University of Michigan pp. 57–94.

Casella and Rauch. 2002. "Anonymous Market and Group Ties in International Trade." *Journal of International Economics* .

Colson, E. 1974. *Tradition and Contract: The Problem of Order*. Aldine Publishing Company.

Compte, O. 1998. "Communication in Repeated Games with Imperfect Private Monitoring." *Econometrica* pp. 597–626.

Conley, T.G. and C.R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *The American Economic Review* 100(1):35–69.

Cook, K.S., R. Hardin and M. Levi. 2007. *Cooperation without Trust?* Vol. 9 Russel Sage Foundation Publications.

Cooter, R. and J. T. Landa. 1984. "Personal Versus Impersonal Trade: The Size of Trading Groups and Contract Law." *International Review of Law and Economics* .

Currarini, S., M.O. Jackson and P. Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica* 77(4):1003–1045.

Dary, D. 1998. *Red blood & black ink: Journalism in the Old West*. Knopf.

DePaulo, B.M., D.A. Kashy, S.E. Kirkendol, M.M. Wyer and J.A. Epstein. 1996. "Lying in Everyday Life." *Journal of Personality and Social Psychology* 70(5):979.

Dixit, A. 2003. "Trade Expansion and Contract Enforcement." *Journal of Political Economy* 111(6):1293–1317.

Dixit, A.K. 2004. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton Univ Pr.

Dreber, A., D. Fudenberg and D.G. Rand. 2010. "Who Cooperates in Repeated Games?" *Working Paper* .

Dunbar, R. 1998. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.

Dunning, T. and L. Harrison. 2010. "Cross-Cutting Cleavages and Ethnic Voting: An Experimental Study of Cousinage in Mali." *American Political Science Review* pp. 1–19.

Durlauf, S.N. 2002. "On The Empirics Of Social Capital." *The Economic Journal* 112(483):F459–F479.

Ellickson, R.C. 1991. *Order Without Law: How Neighbors Settle Disputes*. Harvard Univ Pr.

Ellison, G. 1994. "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching." *The Review of Economic Studies* 61(3):567–588.

Enquist, M. and O. Leimar. 1993. "The Evolution of Cooperation in Mobile Organisms." *Animal Behaviour* 45(4):747–757.

Ensminger, J. 2001. "Reputations, Trust, and the Principal Agent Problem." *Trust in society* 2.

Evans-Pritchard, E.E. 1940. *The Nuer: A Description of the Modes of Livelihood and Political institutions of a Nilotic People*. Clarendon Press Oxford.

Fafchamps, M. 2002. "Spontaneous Market Emergence." *Topics in Theoretical Economics* 2(1):2.

Fearon, J.D. and D.D. Laitin. 1996. "Explaining Interethnic Cooperation." *The American Political Science Review* 90(4):715–735.

Fischer, C.S. 1982. *To Dwell Among Friends: Personal Networks in Town and City*. University of Chicago Press.

Freudenburg, W.R. 1986. "The Density of Acquaintanceship: An Overlooked Variable in Community Research?" *American Journal of Sociology* pp. 27–63.

Fudenberg, D., D. Levine and E. Maskin. 1994. "The Folk Theorem with Imperfect Public Information." *Econometrica: Journal of the Econometric Society* pp. 997–1039.

Fudenberg, D. and D.K. Levine. 1991. "An Approximate Folk Theorem with Imperfect Private Information." *Journal of Economic Theory* 54(1):26–47.

Fudenberg, D. and E. Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica: Journal of the Econometric Society* pp. 533–554.

Fudenberg, D. and J. Tirole. 1991. *Game Theory*. Cambridge: MIT Press.

Gintis, H. 2004. "Modeling Cooperation among Self-Interested Agents: A Critique." *Journal of Socio-Economics* 33(6):695–714.

Glaeser, E.L., D.I. Laibson and B. Sacerdote. 2002. "An Economic Approach to Social Capital." *The Economic Journal* 112:437–458.

Gluckman, M. 1955. "The Peace in the Feud." *Past & Present* (8):1–14.

Goyal, S. 2007. *Connections: An Introduction to the Economics of Networks*. Princeton Univ Pr.

Granovetter, M.S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* pp. 1360–1380.

Greif, A. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 49(4):857–882.

Greif, A. 1993. "Contract Enforceability and Economic Institutions in Early Erade: The Maghribi Traders' Coalition." *The American Economic Review* 83(3):525–548.

Greif, A., P. Milgrom and B.R. Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* pp. 745–776.

Habyarimana, J.P., M. Humphreys, D.N. Posner and J.M. Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. Russell Sage Foundation Publications.

Halaas, D.F. 1981. *Boom Town Newspapers: Journalism on the Rocky Mountain Mining Frontier, 1859-1881*. University of New Mexico Press.

Harbord, D. 2006. "Enforcing Cooperation among Medieval Merchants: The Maghribi Traders Revisited." *working paper* .

Hardin, R. 2002. *Trust and Trustworthiness*. Vol. 4 Russell Sage Foundation Publications.

Hoff, K. and E. Fehr. 2010. "Caste and Punishment: The Legacy of Caste Culture in Norm Enforcement." *Working Paper* .

Hollon, W.E. and C. Crowe. 1974. "Frontier Violence: Another Look." *History: Reviews of New Books* 2(9):221–221.

Horowitz, D.L. 1985. *Ethnic Groups in Conflict*. Univ of California Pr.

Jackson, M., T. Rodríguez Barraquer and X. Tan. forthcoming. "Social capital and social quilts: Network patterns of favor exchange." *American Economic Review* .

Jackson, M.O. 2003. "The Stability and Efficiency of Economic and Social Networks." *Advances in Economic Design* 6:1–62.

Jha, S. 2008. "Trade, Complementarities and Religious Tolerance: Evidence from India." *Working Paper* .

Jha, S. 2009. "A theory of Ethnic Tolerance." *Working Paper* .

Judd, S., M. Kearns and Y. Vorobeychik. 2010. "Behavioral Dynamics and Influence in Networked Coloring and Consensus." *Proceedings of the National Academy of Sciences* 107(34):14978.

Kandori, M. 1992. "Social Norms and Community Enforcement." *The Review of Economic Studies* 59(1):63.

Kandori, M. 2002. "Introduction to Repeated Games with Private Monitoring." *Journal of Economic Theory* 102(1):1–15.

Kandori, M. and H. Matsushima. 1998. "Private Observation, Communication and Collusion." *Econometrica* pp. 627–652.

Karlan, D., M. Mobius, T. Rosenblat and A. Szeidl. 2009. "Trust and Social Collateral*." *Quarterly Journal of Economics* 124(3):1307–1361.

Kluckhohn, C. 1944. "Navaho witchcraft." *Papers. Peabody Museum, Harvard University* .

Kramer, R.M. 1999. "Trust and Distrust in Organizations: Emerging Perspectives, Enduring Questions." *Annual Review of Psychology* 50(1):569–598.

Kreps, D.M. and R. Wilson. 1982. "Sequential Equilibria." *Econometrica: Journal of the Econometric Society* pp. 863–894.

Landa, J.T. 1981. "Theory of the Ethically Homogeneous Middleman Group: An Institutional Alternative to Contract Law, A." *J. Legal Stud.* 10:349.

Laszlo, S. and E. Santor. 2009. "Migration, Social Networks, and Credit: Empirical Evidence from Peru." *The Developing Economies* 47(4):383–409.

Lehrer, E. 1990. "Nash Equilibria of n-Player Repeated Games with Semi-Standard Information." *International Journal of Game Theory* 19(2):191–217.

Lewis, J.L. 2010. "The Initial Stages of Insurgency and Counterinsurgency." *Working Paper* .

Li, J.S. 2003. "Relation-Based Versus Rule-Based Governance: an Explanation of the East Asian Miracle and Asian Crisis." *Review of International Economics* 11(4):651–673.

Lippert, S. and G. Spagnolo. 2011. "Networks of Relations and Word-of-Mouth Communication." *Games and Economic Behavior* 72(1):202–217.

McGrath, R.D. 1987. *Gunfighters, highwaymen, and vigilantes: Violence on the frontier*. Univ of California Pr.

McPherson, M., L. Smith-Lovin and J.M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* pp. 415–444.

Milgrom, P.R., D.C. North et al. 1990. "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs." *Economics & Politics* 2(1):1–23.

Miller, W.I. 1990. *Bloodtaking and Peacemaking: Feud, Law, and Society in Saga Iceland*. University of Chicago press.

Mislove, A.E. 2009. *Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems*. ProQuest.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge Univ Pr.

Prassel, Frank R. 1972. "The Western Peace Officer: A Legacy of Law and Order.".

Rao, N., M. Mobius and T. Rosenblat. 2007. "Social Networks and Vaccination Decisions." *Federal Reserve Bank of Boston, Working Papers* .

Rauch, J. and A. Casella. 1998. "Anonymous Market and Group Ties in Internationale Trade." *Journal of International Economics* 58(1):19–47.

Rubinstein, A. 1979. "Equilibrium in Supergames with the Overtaking Criterion." *Journal of Economic Theory* 21(1):1–9.

Sommerfeld, R.D., H.J. Krambeck, D. Semmann and M. Milinski. 2007. "Gossip as an Alternative for Direct Observation in Games of Indirect Reciprocity." *Proceedings of the National Academy of Sciences* 104(44):17435.

Spencer, R.F. 1972. "The social composition of the North Alaskan whaling crew." *Alliance in Eskimo Society: Proceedings of the American Ethnological Society, 1971, Supplement* pp. 110–131.

Takagi, Y. 2011. "Local Gossip and Inter-Generational Family Transfers: Comparative Political Economy of Insurance Provision." *Working Paper* .

Varshney, A. 2003. *Ethnic Conflict and Civic Life: Hindus and Muslims in India.* New Haven: Yale Univ Pr.

Venkatesh, S.A. 2008. *Gang Leader for a Day.* Penguin Press.

Watts, D.J. and S.H. Strogatz. 1998. "Collective Dynamics of Small-World Networks." *nature* 393(6684):440–442.

Wilson, C., B. Boe, A. Sala, K.P.N. Puttaswamy and B.Y. Zhao. 2009. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems.* Acm pp. 205–218.