Practical Applications of Game Theory in Sociological Research

Rich Lafferty
(9416113)

for Prof. Smith
166-480B, Honours Project

April 17, 2001

Practical Applications of Game Theory in Sociological Research

Rich Lafferty

If one weren't aware of its history, one might think that game theory was about sociology from its beginnings. Where else, after all, would one find the roots of a mathematical methodology which attempts to explain the rational basis of decisions made when people interact, than within the discipline which has as its goal the objective—and often quantitative—study of social mechanisms? Sociologists have long been fond of grand theories, after all, and game theory, by developing methodologies that apply in principle to any interaction and then applying them to particular situationsh (Aumann 1986: 466), seems to offer an attractive basis on which to build them. And game theory is far from a new science; development began as early as 1910, and by 1928—the year in which Von Neumann presented his 'minimax theorem' and had coined the term '*homo economicus*'—it had attracted the attention of economists who saw in it the potential for an umbrella theory for the social sciences (Aumann 1986: 461). By the 1950s—with the discovery of the Prisoner's Dilemma and John Nash's equilibrium concept—it had full status within the disciplines of mathematics and economics (Aumann 1986: 467). In comparison with the lifespan of some sociological grand theories which have come and gone, game theory exhibits remarkable endurance.

Yet despite its potential and pedigree, game theory has not seen much practical application within sociology. Compared to economics—the social science in which game theory has traditionally found its home—sociologists have been slow in discovering the possibilities that a formal mathematical approach to research problems can present. Examining nearly any catalogue which covers both disciplines—journal articles, books in print, dissertation abstracts—one finds that for every sociological publication which addresses game theory, there are ten in economics.[1] Such a discrepancy should not be surprising; there *are* barriers to adopting game theory for sociological research. The most prominent (and, possibly, the easiest to overcome) is that it is mathematically intense; though it may be

[1] Based on a search for 'economics' or 'sociology,' and 'game theory' or 'game theoretical' or 'game theoretic' (or the equivalent with a hyphen), in the *EconLit* and *SocioFILE* periodical indexes, in the *Bowker's Books in Print* catalogue, and in *Dissertation Abstracts International*'s index, from 1996 to 2000.

possible to understand a game-theoretical discussion on a topic without formal mathematical training, it almost certainly would not be possible to *author* one without it.

In this paper, I shall attempt to illustrate the potential for game-theoretical sociological research by examining a selection of articles from the sociological literature which utilize game-theoretical methods to explore areas of research other than game theory itself. While analyzing the approaches of the authors of the selected articles, I will also attempt to identify some typical problems which may arise when using game theory as a sociological research tool, and to suggest recommendations and solutions which avoid or counteract the weaknesses and caveats which appear in sociological game theory as it is being used. At the same time, I will suggest means by which game-theoretical methods can escape their current marginal status and be introduced into the sociological repertoire.

The articles The abovementioned dearth of sociological publications which use game theory was, alas, not dear enough to make straightforward the selection of articles to examine. In order to narrow down the selection, I applied a number of criteria, either explicitly or indirectly through the results of initial searches through the literature. The criteria were:

1. Articles whose results have practical sociological implications are preferred to those whose results are entirely descriptive or theoretical.

2. Recent articles are preferred to those more than ten years old.

3. Articles should represent a diverse set of areas of research within the field of sociology. In particular, articles whose primary conclusions are social rather than economic are preferred.

4. Articles should use game theory as a tool to evaluate hypotheses, build models, etc., rather than using an established result to demonstrate utility of game theory or rational choice theory.

5. Articles in which the application of game-theoretical methods to a problem have reached results which present an alternative to conventional interpretations are preferred over those which confirm results determined through more mainstream methods.

6. Both predictive and descriptive/normative models should be demonstrated by the set of selected articles.

Of the dozen or so articles found which passed the above criteria, five were selected which best illustrated the advantages and difficulties of game theory in a sociological setting, and which originated in a variety of areas of research: industrial relations, medical sociology, social movements, organizations, and criminology. The full text of each article appears in an appendix; in order to conserve space, this paper only summarizes the authors' methods and findings as necessary to discuss its methodological implications, and the reader is encouraged to refer to the original articles in order to fully understand the original authors' intent.

Cliff Brown and Terry Boswell: Strikebreaking or Solidarity in the Great Steel Strike of 1919

In this first article I examine, Cliff Brown and Terry Boswell use a game-theoretical approach to evaluate, alter, and reevaluate a decidedly *non*-game-theoretical theory of labour markets. The Great Steel Strike of 1919 has long been considered a "crucial test" (Brown and Boswell 1995: 1510) of *split labour market theory* (SLM), in which racial conflict is generated by a labour market which is noticeably split between two ethnicities, with the dominant and established race or ethnicity being paid higher than the minority (and often migrant) race(s) or ethnicity(ies) (Brown and Boswell 1995: 1479).

In August 1919, after half a decade of localized labour conflicts, 98% of AFL steelworkers voted in favour of an industry-wide strike. Over 365 000 workers subsequently left their jobs, completely halting production in some plants. By December of the same year, although the strike retained its strength in some cities, it had become apparent that an industry-wide strike could not continue, and the strike was called off (Brown and Boswell 1995: 1496). While there were factors at the national level which took their toll on the strike, including skilled workers' reluctance to leave their jobs and the postwar financial strength of the steel companies, the dominating factor in the strike's failure was at the local level, where the steel industry was able to "undermine the efforts of the workers through violent intimidation and the widespread use of strikebreakers" (Brown and Boswell 1995: 1496; for an analysis contemporary with the strike, see Foster 1920).

The majority of the strikebreakers were black workers imported from the South. According to SLM, minority strikebreaking tends to occur when minority workers are "disproportionately short-term sojourners who would not benefit from long-term struggles" (Brown and Boswell 1995: 1479). That the steel industry was a popular one for migrant black labour and that the majority of strikebreakers in the 1919 strike were black shows how the strike can serve as a "crucial test" for the theory.

Gaps in the split labour market theory The problem with SLM, claim Brown and Boswell, is that the theory makes no account of the converse situation: while SLM tries to explain *lack* of solidarity, it only *assumes* that cases in which solidarity was maintained can be explained by a lack of a split labour market. The existing theory ""fail[s] to explain why and when people overcome an environment of racism to achieve interracial solidarity"

(Brown and Boswell 1995: 1479). Dissatisfied with this assumption, they set as their goal expanding SLM to explicitly account for both possible outcomes.

They intuitively identify four possible limitations of SLM that need to be accounted for in an expanded theory. First, the assumption of *primacy of sojourning migration* needs to be explicitly tested to address the possibility that strikebreaking blacks may not have been sojourners. Second, the *history of the dominant labour movement* needs to be included, as union locals in different cities might have been relatively strong or weak to begin with notwithstanding strikebreaking efforts. Third, since the strike was on a national basis, *government repression* is bound to vary from locale to locale. Lastly, the abovementioned assumption of SLM that *the absence of a split labour market explains interracial solidarity* needs to be explicit in a revised theory of split labour markets (Brown and Boswell 1995: 1481–1483).

Testing the theory Brown and Boswell thus have a hypothesis about split labour markets which they need to test. They accurately identify the strikebreaking problem as one of collective action, and to test it they choose to use Heckathorn's (1989, 1990) model of an iterated mixed sanction system. Heckathorn's model of group-mediated social control is a particularly appropriate choice to use in this instance. It is a generalized model, explicitly designed to be applicable to nearly any problem of collective action with mixed sanctions (Heckathorn 1990a: 367). It is parametric in nature (Heckathorn 1990a: 369–370) enabling it to be easily adapted to particular situations by easily modifying sanction efficacy, monitoring costs, and so forth. It has been expanded by its author to account for stratified systems (Heckathorn 1990a: 370, 1990b). Most importantly, it accounts for the internal control capacity of the group being mobilized and efforts within the group to avoid spillover sanctions (Heckathorn 1990a: 367) where the entire group is punished for one member's infraction.

Heckathorn's theory of collective action models the problem of collective action in terms of an *agent* (imperfectly) capable of monitoring and sanctioning group members for defection, and a *heterogeneous group of actors* in a mixed sanction system in which group members are subject to sanctions directed at individual actors (*individual sanctions*) and those directed at all members of the group (*collective sanctions*), such that the spillover from collective sanctions influences the creation of *compliance norms* in which members of

the group attempt to regulate one another's actions to avoid negative effects of sanctions for others' actions (Heckathorn 1990a: 366–367). In this system, Heckathorn identifies a *free-rider problem*, in which individuals may be tempted to benefit from collective action without incurring the costs of engaging in it (Heckathorn 1990a: 368); in the case of the 1919 strike, receiving the higher wages and improved working conditions which union victory would achieve for all workers, without participating in the strike itself (Brown and Boswell 1995: 1486). At the same time, there exists a *second-order free-rider problem* where actors may be tempted to benefit from the compliance norms created to prevent free riding without incurring the costs of creating and maintaining those norms (Heckathorn 1990a: 368); in the case of the strike, for example, failing to discourage others from working as scabs (Brown and Boswell 1995: 1486).

As such, Heckathorn (1990a: 368–369, 377) identifies six possible pure strategies which an actor might follow in a mixed sanction system. *Full cooperation* involves cooperating at the first level by producing the public good and cooperating at the second level by enforcing compliance norms; in this case, participating in the strike and encouraging others to participate. In *hypocritical cooperation* the actor fails to cooperate at the first level but cooperates at the second level; for instance, encouraging others to picket while crossing the picket line himself. *Private cooperation* involves cooperating at the first level but not at the second—not crossing the picket line, but not encouraging others to do the same. An actor who neither strikes nor encourages others to strike engages in *full defection*, failure to cooperate at either the first or second level.

Considering third-level behaviour—that is, compliance norms about maintaining compliance norms, what Axelrod (1986) labels "meta-norms"—there are two further strategies, which involve active efforts to oppose compliance norms. With *hypocritical opposition*, the actor cooperates at the first level, but at the second level not only fails to cooperate but encourages others not to cooperate at the second level (itself third-level behaviour), while in *full opposition* the actor fails to cooperate at the first and second levels while actively encouraging others not to cooperate at the second level. In the case of the strike, hypocritical opposition might consist of refraining from crossing the picket line while not discouraging others from doing so by picketing, while full opposition might consist of not only crossing the picket line but encouraging others to do so. Oppositional strategies enable actors to actively prevent individually-rational but group-irrational outcomes,

Pareto-suboptimal equilibria in which compliance norms are strong enough that individual costs exceed collective gain (Heckathorn 1990a: 377).

Heckathorn spends an entire article (1990a) deriving payoff functions for all six strategies, so I am unable to account for them all here. It should be noted, though, that for any strategy, the payoff becomes a linear function incorporating actor's opportunities to engage in a particular first- or second-level behaviour (both ego and other), the strength of both the collective and individual sanctions, the efficacy of control and of monitoring, the costs to the individual of cooperating at the first and second level, and the size of the group (Heckathorn 1990a: 373). Actors are assumed to engage in self-interested maximizing behaviour,[2] and each actor makes a decision in turn; in a group of N actors, an actor's decision can only change on the $(N+1)th$ turn, although since each actor is making decisions about whether or not to incur collective sanctions and whether or not to engage in enforcing compliance norms, an actor's *behaviour* can change on any turn (if, for instance, the opportunity to free-ride at the first level ceases to exist) (Heckathorn 1990a: 383).

Brown and Boswell modify the model in two important respects: first, they account for differing opportunities to exert compliant control and differential payoffs for engaging in first-level compliance for the three groups present (that is, the local ethnic majority, the local ethnic minority, and ethnic minority sojourners); and second, they decide that a drop in compliance below 50% diminishes the strength of sanctions and the efficacy of monitoring by 3% to reflect the union's reduced ability to organize over the long term (Brown and Boswell 1995: 1489). Neither of these modifications impact the outcome of Heckathorn's model, which was constructed to account for modifications of this sort (Heckathorn 1990a: 369). Plotting compliance over time, they reach the results shown in Figure 1.

The conclusion is straightforward. There is no question that the strength of the union (that is, its ability to enforce compliance norms) has a significant effect; in fact, while migrant minorities doom the group to eventual failure (0% compliance), groups without migrant minorities *also* fail in the absence of a strong union. From the results of this model,

[2] Note that such behaviour need not require perfect calculative ability. Heckathorn (1996) demonstrates that the outcomes of forward-looking (rational choice), backward-looking (learning) and sideways-looking (observational) approaches converge.

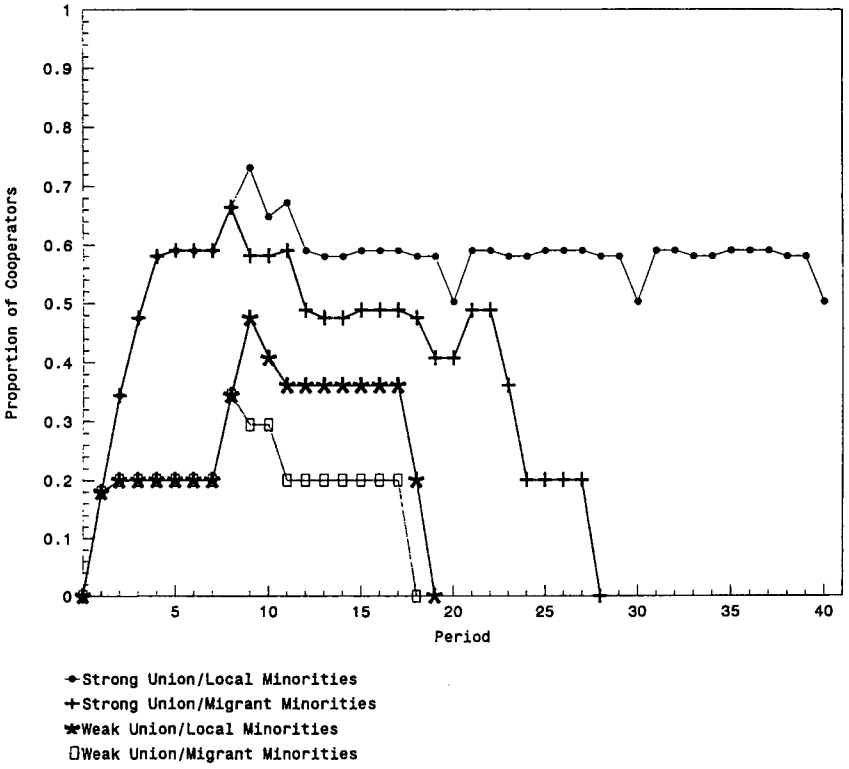

Figure 1: Brown and Boswell's labour market group compliance levels.

they conclude that SLM does not adequately account for conditions in which solidarity is maintained, and modify SLM to expect higher levels of solidarity where a union is comparatively strong *and* recent migrants are few (Brown and Boswell 1995: 1492). But in comparing the expanded SLM to the empirical evidence of the strike[3] they notice that two cities, Cleveland and Wheeling, experienced solidarity despite having either migrant workers or a relatively weak union (Brown and Boswell 1995: 1502).

Brown and Boswell set about identifying other possible independent variables by using qualitative comparative analysis; I omit a description of that methodology due to space considerations.[4] Important to us is that the outcome of their analysis was the identification of a further factor necessary to identify all possible combinations of conditions and outcomes: the repressive nature of the local government (Brown and Boswell 1995: 1505). They modify their collective-action model to account for the increased external sanctions reflective of a repressive local government and recalculate the model's outcomes. Figure 2 illustrates the situation of a pro-union government (with lighter external sanctions), while Figure 3 illustrates the same situation with an anti-union local government (with harsher external sanctions).

It is immediately apparent from these graphs, derived from the application of Heckathorn's model of collective action, that the key to maintaining solidarity—operationalized as an equilibrium at a level of compliance higher than zero percent—is a pro-union local government. In other words, even with migrant minorities, the combination of a strong union and a pro-union government is sufficient to establish labour solidarity (Brown and Boswell 1995: 1508). Brown and Boswell accordingly modify SLM to account for the necessity of a pro-union government to establish racial solidarity (Brown and Boswell 1995: 1509).

Methodological advantages What Brown and Boswell manage to do here—aside from correcting an important inadequacy in the split labour market theory—is to demonstrate how game theory can be productive in sociology without having to write *about* game theory. Their methodology recalls the scientific method: they construct a *hypothesis* on the effects of certain variables on union solidarity; they devise an *experiment* by which

[3] Brown and Boswell (1995) summarize the conditions of the failure of the strike in pp. 1492{1496.

[4] On qualitative comparative analysis, see Regin (1987).

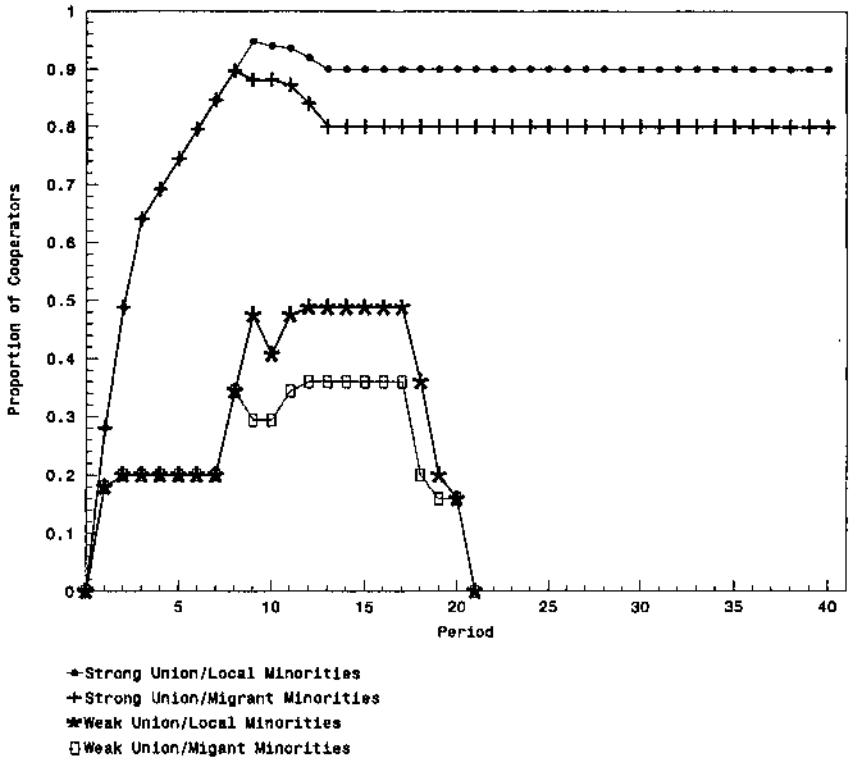

Figure 2: Brown and Boswell's revised model of group compliance levels, pro-union government.

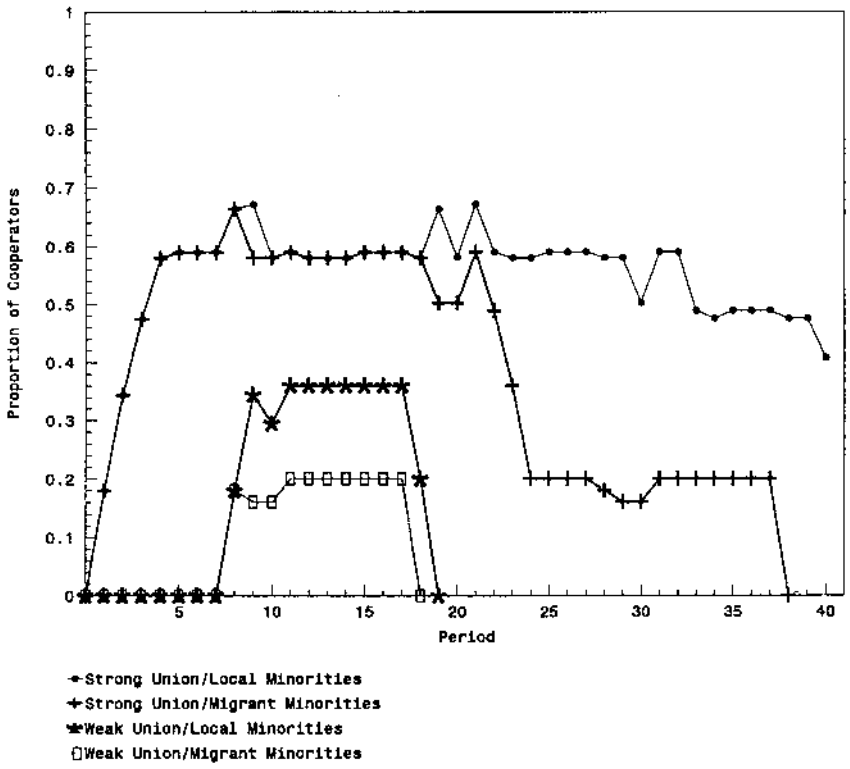

Figure 3: Brown and Boswell's revised model of group compliance levels, anti-union government.

they can test the hypothesis (by observing the output of Heckathorn's collective-action model); they execute the experiment and *evaluate the results*, and when the results of the experiment are inconclusive, they use them as a *feedback loop* to formulate another hypothesis.

What makes Heckathorn's model particularly effective for this analysis is that it was explicitly designed to be general—that is, not a model of a particular situation as in Kahana and Weiss, below, but a general model which others are able to appropriate, easily modify, and apply to a wide variety of situations to produce easily-interpreted results. What contributes to the success and straightforwardness of Brown and Boswell's analysis of SLM is that they are relieved of the responsibility of demonstrating the validity of the model and can instead concentrate on producing results. The mathematical requirements for Brown and Boswell were to ensure that they accounted for the particulars of the situation in ways which would not negatively affect the reliability of the model—means for which Heckathorn provides in (1990b)—and then simply to put representative values out and plot the output, which essentially speaks for itself.

The general principle at work here is essentially an academic *division of labour*. Establishing the reliability and accuracy of the model has required extensive work on the part of Heckathorn, as evidenced by the large body of work he has produced on it (including a few articles not referenced here). By presenting and supporting the formal model separately, Heckathorn enables Brown and Boswell to use it as a tool, applying it to their specific problem domain and drawing conclusions from the results it produces. In other words, the academic roles of game theorist and of labour-market theorist are distinct but their efforts produce a single result, just as, for instance, the owner and worker in the classic industrial division of labour, or the working head of the household and the housewife in the classic household division of labour.

Separating the mathematics of game theory from the application of the theory is standard practice in many of the disciplines in which game theory is applied in a practical manner: in economics, where the abstract economic model is maintained as an ideal to which real-world economies more or less conform; in computer science, where the theories of the game theorists become ideal types in the form of algorithms and libraries; in biology, where theories of evolutionary behaviour are developed independently from the evolutionary traits of a particular species (Binmore 1992: 16–19). The 'soft' social

sciences—sociology and its variations, and political science, for game theory is all but unutilized in anthropology—seem for the most part to have not yet adopted such a division of labour which leaves the mathematics to the mathematicians and social theorists and the practical implications to the rest. Instead, as later articles demonstrate, the burden of establishing the mathematical soundness of a particular approach to a problem is left to those trying to solve the problem in the first place.

Through the judicious application of Heckathorn's formal model of collective action in a mixed sanction system, Brown and Boswell are able to quickly and clearly identify weaknesses in the existing SLM theory of industrial solidarity, test the effects of new variables, and convincingly put forth a modified SLM theory which takes into account their findings and accomodates the empirical cases which failed to conform to the original theory. To be able to generate such important findings and to present them in such a clear manner is a direct result of their having taken advantage of a pre-established game-theoretical model, demonstrating the utility of a division of labour between the game theorists—social scientists who develop abstract, general, and parametric models of social phenomena—and the social researchers who use those tools to address specific questions of sociology.

Kirby Schroeder and Fabio Rojas: A Game Theoretic Model of HIV Transmission

In "A Game Theoretic Model of HIV Transmission: Signaling and Coordination in a Game of Limited Information", Kirby Schroeder and Fabio Rojas attempt to account for the seemingly irrational behaviour of engaging in unprotected sex with strangers who may be HIV-positive. They briefly discuss existing theories for non-adoption of safe sex practices—the "Health Belief" model, the "Communication Perspective", and the "Theory of Reasoned Action" (Schroeder and Rojas 2000: 6)—but claim that none fully account for the interaction between partners in deciding whether or not to practice safe sex, noting that "while these perspectives may provide some accurate *description* of sexual behaviour, they are less able to offer a meaningful *explanation* of it" (Schroeder and Rojas 2000: 3, emphasis in original). Like Brown, they begin with theories which they suspect are incomplete, going so far as to locate the aspects of sexual behaviour not addressed by each. Unlike Brown, however, they discard that body of work, and instead proceed to go about creating a model from scratch.

Their goal, then, is to create a model by which an explanation of risky sexual behaviour can be determined, in which two potential sexual partners negotiate the conditions of the encounter while being able to "keep their own HIV status private" (Schroeder and Rojas 2000: 2), that is, where the potential partners suffer from incomplete information. Unfortunately, as we will see, it is not straightforward to use what are essentially economic methods to model something as non-economic as casual sex.

The Risky-Sex Game Accurately modeling the negotiations leading to a sexual encounter is a formidable task. In order to simplify the encounter enough to be manageable as a game, Schroeder and Rojas initially concentrate on a completely isolated event: a one-time sexual encounter between strangers (Schroeder and Rojas 2000: 11). Their intent is to "focus on the situation where uninfected individuals might encounter an infected individual who values the satisfaction of his own sexual desire over the safety of his partner" (Schroeder and Rojas 2000: 11), although the model itself ends up accounting for all four possible permutations of status.

Schroeder and Rojas set out the conditions for a one-shot casual sexual encounter as follows: Each player in the game knows with certainty his own HIV status as determined

with some unspecified probability by Nature, but cannot know that of the other player. The game explicitly excludes situations in which a player may be HIV-positive but unaware of their status (Schroeder and Rojas 2000: 2). The scope of the game is limited to whether or not these two players will have sex; the option of finding another partner does not exist in the model. The authors note two forms of common knowledge: that the preferences of both HIV-positive and HIV-negative individuals are known to all (Schroeder and Rojas 2000: 16), and that all players believe that the other player has a probability p of being HIV-positive (Schroeder and Rojas 2000: 17).

The game is played as follows (Schroeder and Rojas 2000: 17): Player I offers either protected sex (PS) or unprotected, 'risky' sex (RS). Player II then counter-offers PS or RS, or chooses to not have sex at all (NS), ending the interaction. Player I can then agree to Player II's counter-offer or can choose NS and end the interaction. Lastly, Player II can decide, after Player I's agreement, to either go off and have whatever sort of sex was decided upon, or to end the interaction (NS). After every move, both players update their belief that the other person is HIV-positive. (That the players' beliefs are independently updated after each move implies that p can only be common knowledge, as the authors claim, at the beginning of the game.) The extensive form of the game as appears in the original article is shown in Figure 4. Nature's moves—which determine the HIV status of each player—and the outcomes of each play, are omitted from Figure 4 as they were in the figure provided by Schroeder and Rojas.

Players' Preferences To simplify the determination of the players' preferences, Schroeder and Rojas assume perfect condoms (Schroeder and Rojas 2000: 5)—that is, that sex with a condom does not carry the risk that the condom might break—and that wearing a condom detracts from the sexual experience, *ceteris paribus*.

Schroeder and Rojas explain the preferences of the players as follows: For an HIV-positive player, $RS \quad PS \quad NS$. For an HIV-negative player, $RS \quad PS \quad NS$ if the other player is HIV-negative, but $PS \quad NS \quad RS$ if the other player is HIV-positive (Schroeder and Rojas 2000: 13).

This is a confusing way to state the preferences of an HIV-negative player. As stated, the utility of risky sex fluctuates depending on the HIV status of the other player, which they cannot know. Thus, the utility of risky sex must depend on the player's *belief* about

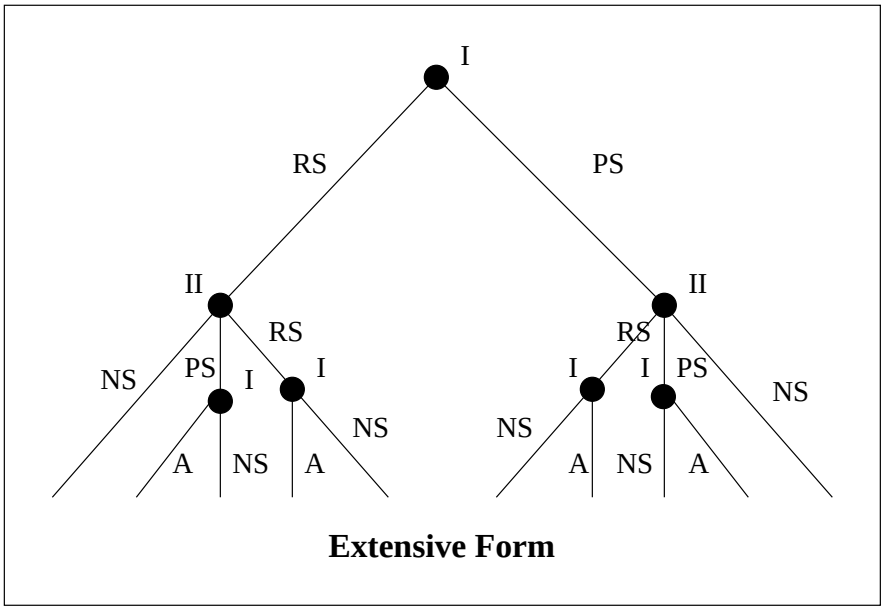

Figure 4: Schroeder and Rojas' (incomplete) risky sex game.

the other player's status. The possibility of intransitivity exists if midway through a game an HIV-negative player changes his opinion on the HIV status of the other player, and by definition a player whose preferences are intransitive is not rational (Binmore 1992: 95). But the preferences as stated seem to make sense; that is, by putting oneself in the position of the HIV-negative player, one can see that safe, condomless sex is the best possible outcome and unsafe, condomless sex is the worst. Is our HIV-negative player irrational, then, or has an error been committed?

The problem with Schroeder and Rojas's preference ordering is that they fail to account for the *totality of outcomes* (Binmore 1992: 95). At any stage in the game, a player can make a *move* of RS, PS, or NS (where the game's concept of 'agreeing' to a proposal is a simple case of making the same move as did the other player in the previous turn). Where the authors slip up is in assuming that a sequence of moves terminating in $[RS; RS]$ gives the outcome RS. While this is true for a sequence of moves terminating $[PS; PS]$ (outcome $[PS, PS]$) or terminating $[any; NS]$ (outcome $[NS, NS]$), there are *two* outcomes for $[RS; RS]$. Which of these two outcomes is reached depends on Nature's moves at the beginning of the game.

In a game between two HIV-positive players or two HIV-negative players, the outcome for $[RS; RS]$ is the best possible, since no condom is used and no player changes status from HIV-negative to HIV-positive. In a game between an HIV-negative player and an HIV-positive player, the HIV-positive player's outcome for $[RS; RS]$ is the same as in the game between two HIV-positive players, but the HIV-negative player's outcome is his least-preferred—being infected. In other words, Schroeder and Rojas appear to have confused moves with outcomes. That their diagram of the extensive form of the game, reproduced as Figure 4, omits Nature's moves (which determine the outcome of a play) and the game's payoffs draws further attention to their omission. They also omit consideration of payoffs, leaving us without any idea of how *much* better non-infectious sex is compared to infectious sex, or how *much* worse sex with a condom is compared to without.

Playing the Game Once Nature makes the first two moves in the game by deciding the HIV status of both players, each player must base his choice upon what he believes the HIV status of the other player to be. As such, Schroeder and Rojas (correctly) treat the game as a *signaling game* (Schroeder and Rojas 2000: 18) in which players manage the signals they send about their status and interpret those they receive about the other player's status.

Schroeder and Rojas contest that the game has both a separating equilibrium and a pooling equilibrium (Schroeder and Rojas 2000: 19). A *separating equilibrium* in a signaling game is one in which the two types of sender send different messages such that the receiver of a message can tell which type the sender is; a *pooling equilibrium* is one in which both types play the same strategy and thus the receiver cannot update his beliefs about the world based on the signals (Morrow 1994: 225). The separating equilibrium is explained by the authors as follows—paraphrased slightly, as I have standardized their terminology:

> If Player I is HIV-positive, he will always offer RS; then, an HIV-positive Player II will counter-offer RS and an HIV-negative Player II will counter-offer PS and Player I will accept either way.
>
> An HIV-negative Player I will offer PS to Player II no matter what he believes Player II's status to ultimately be. Player II, regardless of status, will counter-offer PS.

> Deviations will result in what at least one actor considers a suboptimal outcome. *QED*. (Schroeder and Rojas 2000: 20)

They then proceed to acknowledge that this equilibrium is not separating for Player II (Schroeder and Rojas 2000: 20). But if the equilibrium is not separating for Player II, then Player I does not know anything more about the status of Player II in the third round. Luckily for us, he is not allowed to offer a different kind of sex, but must either agree with Player II or choose NS. Since Player II is always offering PS, Player I will always accept. The authors never formally state what the equilibrium *is*, but from their description we can see that it is $[any; PS; PS]$, with the outcome $[PS, PS]$.

The authors give no explanation for the manner in which the game terminates after two offers. Is it reasonable to assume that if agreement hasn't been reached in two offers, the players will walk away rather than try to coordinate further? A more accurate model seems to be one in which one player choosing NS leads, after one more move by each player, to the outcome PS, which is an improvement for both players regardless of status. In other words, NS, leading to the outcome $[NS, NS]$, is only a credible threat because of the apparently arbitrary limitation in the rules of the game, and is an artifact of the model.

Schroeder and Rojas then go on to explain that for sufficiently high values of p, the game has a pooling equilibrium. Unfortunately they do not provide any hint as to what this value of p might be, how the players' attitudes toward risk influence it, or how the weights of the possible outcomes depend on it or vice-versa. They claim that "the prior belief that [players] are HIV-positive, p, describes the pooling equilibrium where PS is offered which implies that the expected utility of offering PS is more than RS for [Player] II." (Schroeder and Rojas 2000: 21). This introduces the potential for an HIV-negative player to agree to RS if he believes that the probability that the offering player is HIV-negative is large enough. Unfortunately, "large enough" is then explained as a ratio involving the utility of NS and that of RS; since they provide none of the values involved, this simply indicates that some HIV-negative players will have unprotected sex when they believe that the person they are having sex with is HIV-negative without explaining the scenario in which that could arise.

The authors then turn around and note that "the existence of a pooling equilibrium

where [Player II] offers PS depends on how uninfected individuals value not having sex with someone of any type compared to having protected sex with an infected partner" (Schroeder and Rojas 2000: 22). While in reality an HIV-negative person might prefer no sex to having protected sex with an HIV-positive person (from medical risk related to condom breakage or leaks or from moral concerns), *in this game* their preferences are already defined: they will *always* prefer to have protected sex with an infected partner to having no sex at all. It is not clear why the authors introduce the possibility of alternative preference ordering at this point; as with the initial confusion of moves and outcomes, it appears as though they may be allowing elements of the 'real world' to influence their model without explicitly accounting for them. In doing so, the accuracy of the model (even if it were accurate notwithstanding that) suffers, as a new element introduced in passing and after the fact is bound to be handled unsystematically, functioning similar to an uncontrolled independent variable in a quantitative experiment in that its effects on the outcome remains unaccounted for.

Long-term Relationships Schroeder and Rojas also apply their model to long-term relationships. The long-term relationship game is simple: repeated iterations of the casual-sex game described above—including the entire negotiation phase—where outcomes in the future are discounted by a factor of $h < 1$, so as to weight the tradeoff between immediate gratification and long-term gratification (Schroeder and Rojas 2000: 24). The game is repeated indefinitely (Schroeder and Rojas 2000: 24), or, more specifically, repeated until one actor chooses NS.

That this is a questionable model on which to base a long-term relationship should go without saying; certainly there is more to a long-term relationship than repeated casual sexual encounters. But even then, this model does not lend insights when repeated; if the one-shot game separates for Player I, further repetitions are a game of perfect information for at least one player, and if the one-shot game pools, each iteration in the repeated game is identical to the rest, since there are no grounds for either player's beliefs about the other player to change.

According to the equilibrium that the authors describe in the one-shot game, the outcome NS will never be reached, since either player can improve that to PS as long as the other acts rationally. (In fact, the discount should never affect play, since subsequent

rounds will be identical to the first.) Acknowledging that such an outcome does not correspond with the reality of the situation being modeled, they try to account for other situations in which the relationship might end. A player may choose NS "after inferring an undesirable status in the other player" (Schroeder and Rojas 2000: 24). It is left to the imagination as to what an undesirable status might be, although they later note that "the relationship ends if an HIV-negative individual suspects that his or her partner is HIV-positive" (Schroeder and Rojas 2000: 24). This is plainly inconsistent with their model. An HIV-negative player *prefers* protected sex with an HIV-positive player to no sex at all, and external factors such as finding another partner—and the costs associated with such factors—do not enter the model at all.

From here, the authors continue to introduce externalities for which the model does not account. For instance, they note that "in a long-term monogamous relationship, two HIV-negative individuals typically discontinue condom use as they solidify the terms of their relationship and develops a shared level of trust" (Schroeder and Rojas 2000: 6), and that "it is also likely that insisting on condom use signals a lack of trust, a key ingredient in the subjective value of the sexual relationship ... thus one actor's offer of PS in the repeated game may lead the other actor to end the relationship" (Schroeder and Rojas 2000: 24). Now, this may be true, but it is not the outcome of *this game*, which places no value on trust or fidelity at all. One is led to suspect that the authors have by now realized that their model fails to account for the necessary factors; were that the case, then the problems once identified should be addressed within the game itself, rather than being introduced in passing as special cases in the repeated version of the game!

As their analysis proceeds they introduce still more factors external to the model; distrust of HIV-positive partners (Schroeder and Rojas 2000: 14), questioning one's own HIV status, altruistic condom use by HIV-positive partners with HIV-negative partners (Schroeder and Rojas 2000: 16), avoiding confrontation, extramarital sex and obtaining new partners, and revealing one's status to one's partner (Schroeder and Rojas 2000: 25). At this point the initial model is all but abandoned, and while they may indeed have insights into condom use in long-term relationships, they remain unsupported as the game in which they imputedly take place is no longer the same game.

Evaluation "This game may possess an equilibrium that is not described here," conclude the authors, but they "have reached [their] goal—to develop a descriptive model of condom use" (Schroeder and Rojas 2000: 23). (They have by this time apparently abandoned the goal of an explanatory model to which they referred twenty pages previous.) I cannot see where such a goal was reached. Schroeder and Rojas' article primarily suffers from a lack of rigour: arbitrary restrictions with powerful effects on the outcome of the game, conclusions reached based on probabilities never specified, and preferences modified—as if they were dependent variables—based on the outcomes of a model which relied on them being fixed.

Throughout the article one gets the impression that the authors are attempting to build a model whose *outcomes* mirror those found in empirical studies. Even before the game is introduced, the authors are discussing theories and experiments (Schroeder and Rojas 2000: 10–13), and before introducing the game note that the success of the model should be judged by how well it reflects conclusions reached by earlier theories (Schroeder and Rojas 2000: 13). The fallacy here is that reaching a particular conclusion does not mean that the model is accurate; there may exist multiple models that reach the same outcome for a given input but which fail to accurately reflect the chain of events from input to output. A liberal application of Occam's Razor would suggest that their four-move sexual negotiation is not a representative predictor of the outcome of a sexual encounter, and it is difficult to come away from the article without the impression that they have realized that themselves when trying to apply the model to the long-term relationship. Rather than modeling the *processes* proposed in the existing theories of sexual behaviour which they discuss, they only ensure that they reach the same outcome. As a result, instead of generating insights as to where existing comprehensive theories of risky sexual behaviour fail, they produce only a simple and unsystematic model which attempts to account for one very specific (and unrealistic) scenario.

Schroeder and Rojas's simple game-theoretical model of sexual behaviour seems to be a questionable approach from the beginning. Not only are costs involving finding new partners not considered, but the preferences upon which the model is based are those of a distant onlooker; while many would arrive at the same preferences if asked on the street, I strongly suspect that one's preferences might be ordered differently when opportunity presents itself in the one-shot game—especially considering the potential influence of intox-

icants and hormones on the decision-making process—and that a game of repeated casual sexual encounters is equally inappropriate to model a long-term intendedly-monogamous relationship. In combination with a seemingly arbitrary model of sexual negotiation, it remains unclear how Schroeder and Rojas's model can contribute to our understanding of the factors leading to the employment of safe-sex practices. It does however illustrate the importance of rigour in the application of game-theoretical methods to social research; had the authors refrained from introducing externalities while trying to make their model fit empirical observation, they may have paid closer attention to the appropriateness of the model in the first place.

Upon reaching the end of Schroeder and Rojas' work, one is left wondering what has been accomplished. The potential for practical insight was high; there were, after all, a number of existing theories each of which, according to the authors, "represents an attempt to understand high-risk behaviour, but ... also fails to address the fundamental problems of trust, deception, and behavioural problems that are resolved in the game theoretic model" (Schroeder and Rojas 2000: 6). But even if we ignore the shortcomings of their model, we still end up further from an understanding of high-risk behaviour than we would be with those existing theories, and with no way to integrate them to salvage this model. While the inconclusiveness of Schroeder and Rojas's article is primarily a result of the weaknesses of their model, I would suggest that, had they attempted to evaluate the weaknesses they identify in the existing corpus directly—by using them as the basis for a game-theoretic model, and identifying via the model the areas in which the theory is unsatisfactory—that results more directly applicable to the study of sexual behaviour could have been extracted.

John Ginkel and Alastair Smith: So You Say You Want A Revolution: A game theoretic explanation of revolution in repressive regimes

In "So You Say You Want A Revolution: A game theoretic explanation of revolution in repressive regimes," John Ginkel and Alastair Smith attempt to create a game-theoretic model of the information and decision processes leading to revolution. As we will see, their attempt appears reliable and complete, if preliminary; what makes it interesting are the similarities it shares with Schroeder and Rojas' considerably less successful HIV model discussed above. Like Schroeder and Rojas, Ginkel and Smith have set about creating their model to rectify the absence of certain elements in the conventional literature, choose to use a signaling game as the basis for their model, include a number of variables which are very difficult to quantify, and include only a small proportion of the factors that might be considered relevant in the situation they are modeling. Yet one is left with the impression that Ginkel and Smith's model works—that while it is unquestionably a simplification, it is a useful framework for analyzing real events. As such, we can use it to suggest practices to be encouraged in game-theoretical sociological research.

The Revolution Game Ginkel and Smith's model of the revolutionary process involves three players: a government who may or may not be overthrown, a group of dissidents who may or may not try to mobilize the public, and a 'mob' representing the public who may or may not heed the dissidents' call for mobilization (Ginkel and Smith 1999: 292). The dissidents differ from the mob in that they are willing to protest; as "marginal citizens under the government regime" they do not worry about a loss of status (Ginkel and Smith 1999: 293). The mob, on the other hand, will not revolt on its own, but relies on the dissidents to tell them when to protest. However, the mob will not necessarily follow the dissidents if they think the revolution will fail (Ginkel and Smith 1999: 293).

The problems that the game accounts for involve the *movement of information*. The government has a type θ which represents the probability of withstanding revolution; that is, high θ means a strong government (Ginkel and Smith 1999: 293). But the dissidents cannot observe θ directly; instead, they possess noisy information n from which they must deduce the government's type. Since the regime is repressive, there are limits on the flow of information, so the mob has little to no idea about the government's type; thus, the dissidents' knowledge of θ is between that of the government (who may not know their own type but has more information on it than the dissidents) and the mob (Ginkel and Smith

1999: 294). However, the nature of the repressive regime leads citizens to withhold their own opinions, so neither the government nor the dissidents know how the mob will respond to a call to revolt (Ginkel and Smith 1999: 294); the dissidents must thus decide whether or not to begin the process of revolution based on their *beliefs* about the government's type and the mob's opinion.

The game proceeds as follows (Ginkel and Smith 1999: 294):

1. Nature determines the government's type, θ 2 $[0, 1]$, and provides the dissidents with noisy information n about the government's type.
2. Given its type, the government may offer an accommodation to its citizens, a 2 $[0, 1]$.
3. The dissidents decide whether to mobilize or accept the accommodation. Without mobilization, the game ends and the accommodation is implemented.
4. If the dissidents mobilize, then the mob decides to participate with some probability γ. If the mob fails to participate, the game ends with the accommodation a being accepted.
5. If revolution occurs, it is successful with probability 1 θ; if it fails, the government withdraws its offer of accommodation.

Players' beliefs The outstanding difference between this game and Schroeder and Rojas's risky-sex game is that each player only moves once; there is no opportunity for a particular player's beliefs to change between moves, thus eliminating the situation in which the outcome of the game varies when the number of negotiation rounds (*i.e.*, moves) is changed. The dissident's beliefs about the government and mob are updated prior to having made a move, and while they account for a from within the game, there is no restriction on how they obtained n on which they initially base their beliefs on the government's type θ. Similarly, while the mob uses a and the dissidents' move in deciding whether or not to revolt, their beliefs about θ and the reliability of the dissidents is established before they move and irrelevant for the updating of beliefs after they move.

Thus each player has to update his beliefs exactly once. Ginkel and Smith predictably update beliefs with Bayes's Theorem (as Schroeder and Rojas did), taking into account

an extensive set of variables (as Schroeder and Rojas *didn't*). Since Bayes's Theorem is essentially a practical formulation of the definition of conditional probability (Morrow 1994: 164), increasing the number of variables—in Bayesian terms, 'states of the world' (Morrow 1994: 164)—allows players' modeled beliefs to better represent the situation being modeled. Aside from the terms already introduced, Ginkel and Smith (1999: 297–299) include the following variables: the danger of violence, the cost to the dissidents of exposure (*i.e.*, to punishment if the revolution fails), the dissident's and the mob's reward from having toppled the government (each independently), and the government's reward from maintaining office. By explicitly listing the variables (Ginkel and Smith 1999: 298) they provide themselves a safeguard against inadvertently introducing extraneous details into their model, as any other factor introduced will be conspicuously missing from the variable list. The outcomes of the model are straightforward: a successful revolution, a failed revolution attempt, the mob's decision not to mobilize upon the dissident's call, and the dissident's decision not to mobilize at all (Ginkel and Smith 1999: 298).

Equilibria and observations The contributions of the model come not from observing outcomes for a particular set of inputs—since, after all, these quantized variables cannot be measured in real situations—but from the interactions between the parties during the playing of the game, *i.e.*, at Bayesian equilibria (Ginkel and Smith 1999: 299). Ginkel and Smith break down their observations by move, and explain the *implications* of the model in the body of their article, relegating their derivation and some less interesting properties to an Appendix. For the mob, they observe that a mob who has more to gain from change will be more likely to rebel (Ginkel and Smith 1999: 299), and that a stronger government will make the mob less likely to rebel *despite their inability to directly observe the government's type* (Ginkel and Smith 1999: 230); for dissidents, they observe that dissidents are more likely to mobilize if the government is weak, and when the mob is likely to follow (Ginkel and Smith 1999: 301). They also observe that their model captures the concept of credibility: that the likelihood that the mob will follow the dissidents is inversely related to the dissidents' cost of exposure (Ginkel and Smith 1999: 301). They also note that accommodations from the government will be less frequent than naïvely expected, in that while they reduce the likelihood of revolution, they serve as particularly noise-free information about they government's type (Ginkel and Smith 1999: 303).

Most importantly, though, they reach their goal of introducing new factors into existing theories of revolutionary behavior: their model offers an explanation of the manner in which repressive regimes last in the face of public discontent. Kuran (1995) hypothesizes that public preference falsification in the Soviet Union was essential both for its tenure and for the rapidity of its fall (Ginkel and Smith 1999: 303). Ginkel and Smith note that in order for the game to be played all the way to revolution, there must be enough information introduced into the system to allow the dissidents' and the mob's beliefs to update such that revolting is rational. With pervasive preference falsification, the dissidents never receive new information on the willingness of the mob to revolt, and therefore will never protest. Without protest, the mob continues to believe that the government is strong—a situation Ginkel and Smith characterize as "informationally frozen" (Ginkel and Smith 1999: 303). Without new inputs, conditional probabilities never change, and no-one acts. Similarly, a new event *can* give the dissidents information that leads to protest, triggering a chain of mob and government reactions that push *more* new information into the system very rapidly, leading to what appears to be the sudden fall of the regime.

Evaluation While the results reached by Ginkel and Smith are not new, their confirmation of Kuran's hypothesis and their observations about information flow are certainly convincing. What makes this study a convincing one is that the authors support the accuracy of their model with *empirical evidence*. Half of the article is comprised of applications of the model to the Czech Velvet Revolution and Chinese Tiananmen Square protests of 1989. In the case of the Velvet Revolution, they note that increasingly generous accommodations from the government sent signals to dissidents that the government was weak, while crackdowns on vocal dissidents sent signals to the mob that the dissidents were credible (since they exposed themselves to significant risk in calling attention to themselves), culminating in a successful call to revolution (Ginkel and Smith 1999: 306). In the Tiananmen case, an information freeze effectively limited the dissident's knowledge of the state's type, and the precipitating factor—the death of Hu Yaobang, a student-friendly Communist party member—provided a symbol around which to rally, but did not provide any new information with which to update beliefs (Ginkel and Smith 1999: 308). While draconian crackdowns on dissident activity made dissidents credible in the mob's eyes, the protestors misevaluated the strength of the state, which was able to withstand the protests (Ginkel and Smith 1999: 310). While two empirical cases are insufficient to

conclusively determine the accuracy of the model—which the authors admit (Ginkel and Smith 1999: 311)—they do provide an acknowledgment of the model's validity which is more reliable than, for instance, Schroeder and Rojas's 'common-sense' observations.

Ginkel and Smith also rely on what might be called 'logical evidence' in determining the accuracy of their model. These tests *do* rely on 'common-sense' observations, but only insofar as to ensure that the development of the model is not headed in a blatantly wrong direction. For example, they observe that a mob which desires change most will be most likely to rebel (Ginkel and Smith 1999: 299), that a mob who thinks the government is strong will be less likely to rebel (Ginkel and Smith 1999: 300), and that dissidents are more likely to mobilize against a weak government (Ginkel and Smith 1999: 301), all of which suggest that their model is on the right track. In the Appendix, in which they demonstrate the application of Bayes's Theorem, they reach a number of lemmas that serve a similar purpose in verifying the model's accuracy, but which are not as obvious on the surface: that dissidents will not protest if the probability of the mob mobilizing is 0 (Ginkel and Smith 1999: 314), that a regime will not offer an accommodation that increases the probability of revolution (Ginkel and Smith 1999: 315), and that stronger states offer smaller accommodations (Ginkel and Smith 1999: 315). These strictly mathematical relations offer evidence that the model will not suffer from spurious inaccuracies. (I suspect, for instance, that had Schroeder and Rojas rigourously computed their repeated risky-sex game, they would have found it impossible to construct similar lemmas for their claims of the effects of trust and fidelity on long-term relationships.) By examining logical evidence, Ginkel and Smith ensure that their model follows the path that the real events it models follow, and by examining empirical evidence, they lend authority to the conclusions they draw from the model.

The presentation of the model also contributes to its practicality. Like Brown and Boswell, Ginkel and Smith acknowledge that the social researchers who may wish to profit from their conclusions need not understand each step of the mathematics by which that conclusion was reached; where Brown and Boswell drew on previous work by Heckathorn, Ginkel and Smith hide the details from the casual observer in an appendix, such that those interested in verifying the mathematical accuracy of their model can still do so. The benefits of separating derivation from presentation are clear when this article is compared to Schroeder and Rojas's; one possibility is that the lack of rigour which plagues Schroeder

and Rojas's work was the result of a misguided attemt to keep their presentation accessible.

Finally, Ginkel and Smith's model succeeds because of the game itself. It is unquestionably a simplification, but they simplify without endangering its premise. I suspect that they accomplish this because their simplifications involve only omission. In comparison, Schroeder and Rojas simplify with constraints; by limiting the sex negotiations to their four-step game, they *introduce* elements for the sake of simplicity, which ends up introducing artificialities. By limiting themselves to omission, Ginkel and Smith avoid artificialities; in particular, limiting the game to a single move per player, and putting ranges on the variables which strongly affect outcomes allows the model to function within a well-defined domain, eliminating extreme (inaccurate) results. Omission-based strategies are not free from inaccuracy. That things have to be omitted means that the phenomenon being modeled is at a much higher level of complexity than the model represents. While the revolution game performs admirably in this respect, there are undoubtedly complications incompatible with the model which could influence the results: for instance, what if a government, after offering an accommodation and experiencing dissident (but not mob) protests, chooses to withhold the accommodation from the mob? What if two groups of dissidents clash on whether or not to mobilize? There appears to be a ripe opportunity to merge this signaling-game model with a model that accounts for second- and third-level effects in collective action, as did Brown and Boswell with Heckathorn's. Nonetheless, the combination of rigour, effective simplification and empirical testing lets Ginkel and Smith's revolution model make important practical contributions to the non-game-theoretical literature.

Nava Kahana and Avi Weiss: Absenteeism: A Comparison of Incentives in Alternative Organizations"

In "Absenteeism: A Comparison of Incentives in Alternative Organizations," Nava Kahana and Avi Weiss identify a moral hazard in the organization of work in that workers have the opportunity to be claim sickness and be absent from work when they are not sick (Kahana and Weiss 1992: 573). They identify the problem as one of the employer's incomplete capacity to monitor the employee—that there is an unobservable probability that the worker is sick, but that only his absence, and not his sickness, can be observed (Kahana and Weiss 1992: 574)—and that that situation can lead to strategic behavior on the part of the worker.

In particular, they intend to "analyze the incentives" that workers have to not work when they are healthy (Kahana and Weiss 1992: 574), a problem well-suited to the application of game theory. They do so by exhaustively developing a simple two-person, single-period model of absenteeism in a labor-managed firm—a cooperative that divides its income equally between members such as a Kibbutz (Kahana and Weiss 1992: 574)—and then by expanding that basic model to account for a variety of more practical situations. Beyond basic game theory the model is self-contained in that it draws little on previous game-theoretical work; while with Schroeder and Rojas earlier we encountered an example of the risks of proceeding in such a manner, Kahana and Weiss are able, for the most part, to successfully devise a general model and refine, permutate, and evaluate it within one paper. Doing so has led to an exceptionally dense exposition; they explain in a page what other authors mentioned in this report would take five or six pages to cover. As such, it is impractical to enter into full detail of their expanded models here. I will instead explain the basis of their simple model and the nature and implications of their extended models while omitting much of the process.

The Absenteeism Game Kahana and Weiss (1992, footnote 2) explicitly note that in exploring this topic they had in mind "a Kibbutz-type setting where, for idealogical [*sic*] reasons, income is divided equally, independent of how often one is 'sick'." Thus they begin by examining the simplest case of absenteeism, a two-person cooperative labour-managed firm (LMF) which divides its income equally between its members. Each member works ($l_i = 1$) or does not work ($l_i = 0$), and total production is equal to the square root of the productive abilities of each worker on the job, $y = L^{0.5}$ where $L = l_1 + l_2$ (Kahana and Weiss

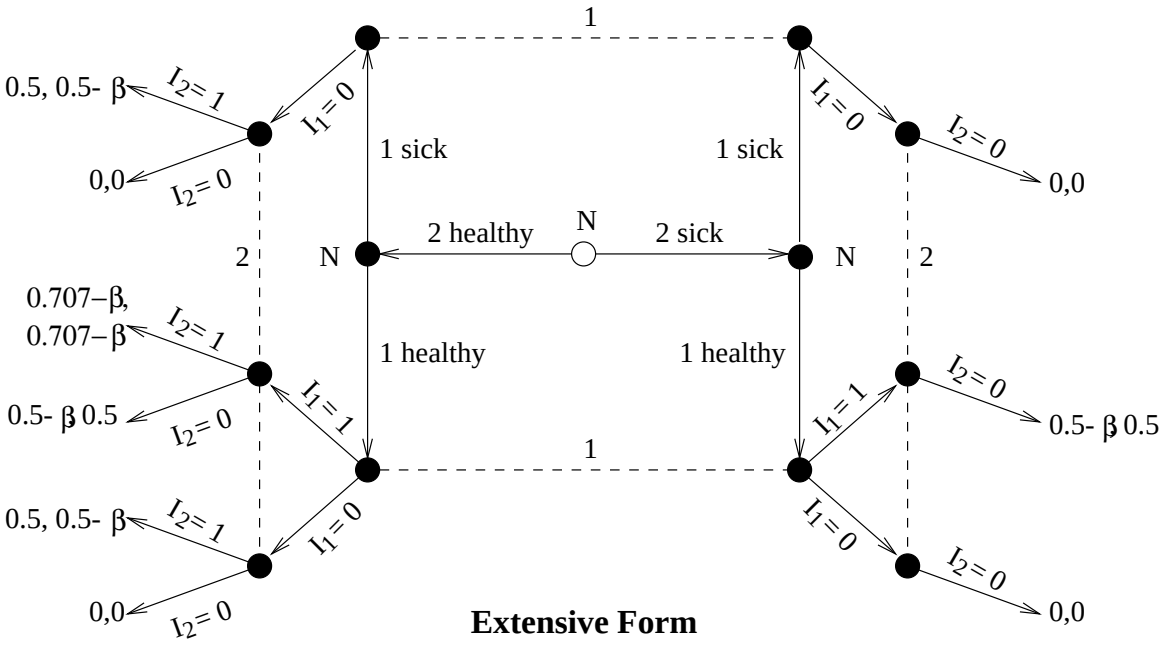

Figure 5: Kahana and Weiss's simple absenteeism game.

1992: 575). Thus, if one worker works, total production equals 1, and if both work, total production is equal to *less* than two, taking into account the nonlinearity of productivity. Nature makes the first two moves of the game, deciding whether or not each player is sick or healthy (and is explicitly accounted for, unlike in Schroeder and Rojas); then each player decides independently whether or not, if healthy, he is going to work. A sick player never works, and players cannot communicate beforehand. Payoffs for a player are equal to his income minus his disutility from working (notated β), or $L^{0.5}/2 \quad \beta l_i$ (Kahana and Weiss 1992: 575). The extensive form of the game is shown in Figure 5.

The extensive form of the game is representative of the thoroughness of their exposition; every value is accounted for, as should be expected but has previously not always been found. They demonstrate the solutions to the game as a tradeoff between a player's disutility from work and the probability of the other player being sick (Kahana and Weiss 1992: 577); the solution space is shown in Figure 6. As the figure shows, there are four regions; in the upper and lower areas, not working and always working dominate all other strategies because of the high and low values of β, respectively. The interesting areas

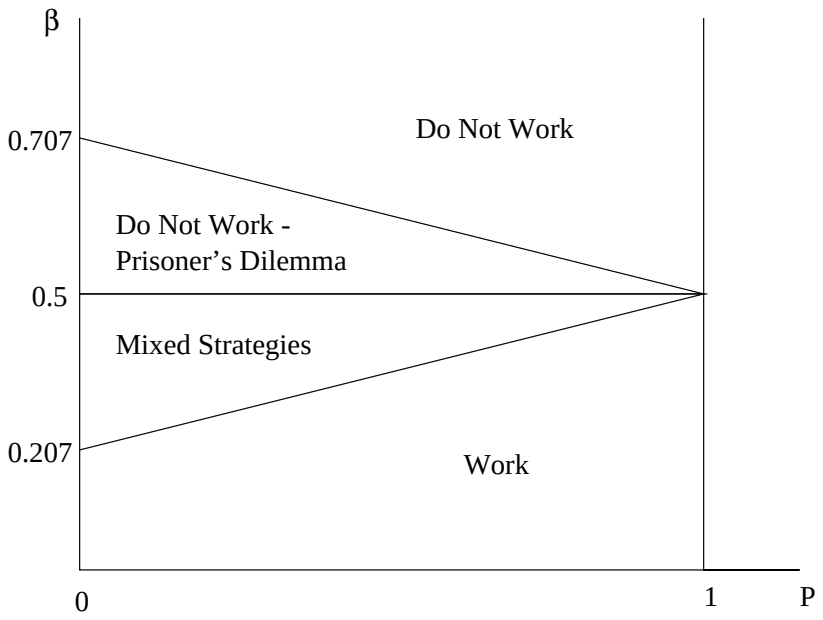

Figure 6: Solution space to the Kahana and Weiss's simple absenteeism game.

of the solution space are those in the middle; in the region above centre an equilibrium occurs when neither worker works, but since that equilibrium is Pareto-suboptimal to both working (that is, if both were to work the outcome would improve for both players), the outcome is a Prisoner's Dilemma, and in the region below centre, there are two Nash equilibria—strategies in which no player can improve his outcome without requiring action on the part of other players—in which one player works and the other does not, and the outcome is a Battle of the Sexes (Kahana and Weiss 1992: 575). Kahana and Weiss refrain from examining the implications of these findings, a decision reflective of the efficiency with which they expose their model and its implications throughout—after all, the purpose of the simple model is to be built upon further, not to deliver immediate results, so the outcomes of this building-block model are of little practical use.

Expanding the Model Having created a model on which to build, Kahana and Weiss take us through a whirlwind tour of modifications to the model which *do* provide practical insights. The first and most substantial modification is in nite repetition of the simple game (Kahana and Weiss 1992: 578). Again, they maintain their efficient exposition,

this time by limiting themselves to three cooperative alternative outcomes: never work; alternate, such that only one player works on a given day; or always work when able. This reduction of the set of outcomes follows from the observation that, since utility is linear, any other alternative could not be undominated since it would be a linear combination of the three outcomes listed (Kahana and Weiss 1992: 578). The outcomes correspond to regions of the solution set in Figure 6, and are for the most part self-enforcing. In the area in which self-enforcing solutions are not Pareto-optimal, they develop a trigger mechanism which would determine the optimum punishment for an absence of a particular length, and then immediately discard the trigger for certain levels of β, recommending the Pareto-suboptimal alternative of alternating days in order to avoid unnecessary punishment resulting from being absent from sickness long enough to trigger it (Kahana and Weiss 1992: 582), further evidence that few details have escaped consideration.

The simple model is then modified to account for pro t-maximizing rms (PMFs) in which workers are paid on an hourly or daily basis, reducing the problem for each player to the choice of "how many hours to work and how much leisure to consume" (Kahana and Weiss 1992: 583). Basing wage on expected marginal product, they demonstrate that at equilibrium the wage rate must equal β to avoid extreme participation rates (Kahana and Weiss 1992: 584)—in other words, that the problem of absenteeism in PMFs is one of inefficiencies resulting from workers working too much, a reflection of the reduction in output per player when both players choose to work; they further note that the PMF does not suffer from the LMF's problems of underparticipation (Kahana and Weiss 1992: 586), an important result in itself.

But the authors do not stop there; they go on to introduce effort into the game, in which a worker in an LMF, having chosen to go to work, can then shirk on the job (Kahana and Weiss 1992: 586). Again representative of their efficient exposition, the implications of that change are simply expressed as differences with the initial solution shown in Figure 6: areas occur in the region where players previously always showed up to work but now play mixed strategies (Kahana and Weiss 1992: 589). They immediately move to the implications of a repeated game with shirking, and reduce it to the earlier problem of choosing between the risk of an easily-triggered trigger mechanism or the low risk but pareto-suboptimality of a self-enforcing alternating solution (Kahana and Weiss 1992: 590). But this still isn't enough, as they then proceed to analyze the initial game

with more than two players. They again express their results as a comparison with their original solution space (figure 6), noting that the only change is that the area in which players choose to always work becomes smaller as the number of players increase (Kahana and Weiss 1992: 590), and they discuss but do not get distracted by the implications of an n-person analog to the Prisoner's Dilemma (Kahana and Weiss 1992: 592)s.

Evaluation As mentioned above, the most attention-grabbing aspect of this study—and what makes it a success—is the efficiency with which the authors are able to build, expand, evaluate, and analyze their models, which comes across in the article itself as *density*. Unlike some of the other authors' works discussed above, Kahana and Weiss resist being sidetracked by *gedankenexperiments* on elements which are not included in their models. They account for assumptions they make in such a way that the reader is satisfied that they have been addressed, and in such a way that the reader understands the implication of the assumption, without distracting from the model that makes the assumption in the first place; for example, they note in a footnote on p. 574 that the assumption of an egalitarian division of profits might seem problematic because each worker's effort is observable, at least in terms of when one shows up at work, but then address in footnotes on pp. 583 and 586 the implications of paying LMF workers based on effort, rather than in an egalitarian division. In the worst case, the reader is left wondering what would happen if a different assumption had been made (for instance, where they assume that there is no communication between workers prior to arrival at work (Kahana and Weiss 1992: 576)), but never left wondering if a miscalculated assumption might have negative effects on the accuracy of the outcome. As well, by expressing outcomes of the later games in terms of the solution space of the simple game, they are able to convey the *differences* resulting from the modifications, rather than forcing the reader to understand each outcome independently.

The result of the authors' density of exposition is that they manage to avoid the problem that Schroeder and Rojas encountered, in that they are not limited to an impractical model because of space considerations. Instead, they have done the equivalent to what Brown did in using Heckathorn's model of collective action; they have taken a general if abstract model (the one-shot LMF game) and applied it to practical scenarios (the repeated games, the PMF games, and the more complex shirking and n-person games), making up

for the lack of suitable abstract models in the literature with a short but complete one of their own. This approach is certainly more successful than only covering a simple model as some of the other authors discussed here have done, but it also has its disadvantages. The primary disadvantage of this method is straightforward—sociologists with the ability to develop and expand a model in this manner will be relatively rare compared to those who are able to take a developed model and apply it. Nonetheless it remains important that the possibility of undivided execution exists. The second disadvantage is that the density of exposition brings with it problems of accessibility; Kahana and Weiss simply do not have the time to explain the theoretical elements behind the model, and since, unlike in Brown, understanding the theoretical elements is a necessary precondition to understanding the applied model, the audience for this paper is necessarily limited to those with a suitable level of understanding of game theory.

At the end of their article, Kahana and Weiss note that their model has empirical implications and suggest some testable hypotheses for further research, and it is here where this paper leaves something to be desired in terms of *practical* sociology. While the authors use logical tests throughout to ensure that their model represents what is being modeled—for example, noting that the results in Figure 6 suggest that as the probability that the other worker will not be there increases, it becomes more important for this worker to show up (Kahana and Weiss 1992: 578)—we do not know at the end of the paper whether or not *any* of the implications the models suggested actually hold true, simply because we do not know if the models themselves hold true. Logical tests tend to rely on common sense, which while at times useful risks overlooking cases in which common sense does not prevail; for example, Taylor's innovative correlation between gun control measures and costs to the victim, discussed below. Empirical tests—as simple as checking to see that the predictions a model makes come true—*must* be the final step in determining the practical applicability of a game-theoretic model.

Robert Taylor: A Game-Theoretic Model of Gun Control

In "A Game Theoretic Model of Gun Control" (Taylor 1995), Robert Taylor attempts to model the effects of gun control on the utility of the innocent citizen, the potential victim of violent crime. Specifically, he sets out to provide theoretical support for Polsby's (1993) contention, published in the Libertarian magazine *Reason*, that "even if univesral disarmament is optimal, gun control measures that merely move us in that direction may *reduce* social welfare." (Taylor 1995: 270, emphasis mine). Empirical studies having been inconclusive, he argues, a more theoretical approach and a simplified model may turn up insights previously passed over.

Taylor, then, begins with a well-defined hypothesis to test, and the basic outline of an experiment with which to test it: he intends to develop a normative game-theoretical model of gun control which demonstrates contradictions to two implicit assumptions that "most arguments for gun control make:" that "universal disarmament is an optimal state," and that "gun control, by moving us toward this optimal state, will necessarily increase welfare" (Taylor 1995: 270), and which can be used to determine the direction which gun-control policy should take. Unlike Brown and Boswell, Taylor chooses to start from scratch on his model.

Taylor's Games Taylor borrows from the hard sciences in his methodology: first, generate a simple model based on acknowledged assumptions, and then from that develop a complex model which accounts for as many of the assumptions made in the first model as is feasible. This approach is somewhat foreign to sociology, with its love of grand theories and distaste for intentional simplification, but is well-established in hard sciences (*cf.* the hoary old physicist-biologist joke which begins, and often ends, "Imagine a perfectly spherical elephant in a vacuum. . . "), and for the purpose of making policy, being able to examine the effects of the element being controlled by policy in isolation can be useful—assuming that the variables in question can, in fact, be isolated. At the same time, he assumes the burden not only to ensure that his conclusions logically follow from the model, but that the model itself is sound and consistent.

Taylor's basic model is a one-shot simultaneous-move game played between Player I, a potential victim who may or may not be armed, and Player II, a criminal who may or may not attack Player I. The game is played as illustrated in Figure 7.

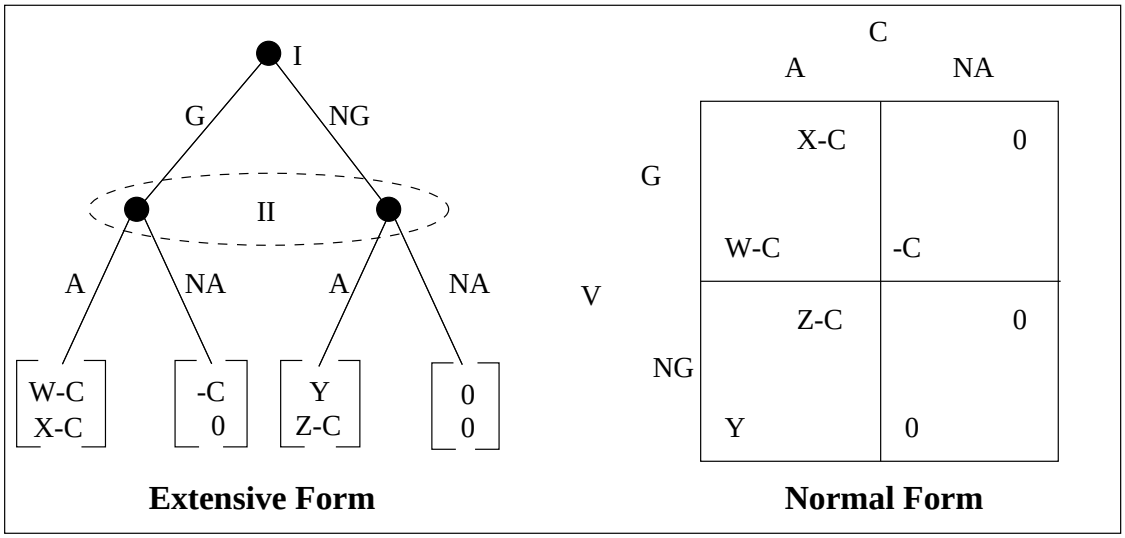

Figure 7: Taylor's basic gun-control game.

Simce Player I and Player II move simultaneously, it is a game of imperfect information (Binmore 1992: 501); the extensive form shown in Figure 7(a) could just as easily show the criminal as having the first move. Player I and Player II are assumed to be risk-neutral to simplify the calculation of payoffs. C in Figure 7 represents the cost of obtaining a gun, and is the sum of a constant, C , representing the price of a gun, and a variable, R, representing additional costs imposed by gun-control measures (Taylor 1995: 271). Most importantly, varying R changes the nature of game's Nash equilibria; the game is not a particular type (Prisoner's Dilemma, Chicken, etc.) but one whose equilibria shift according to the R of a particular policy choice. The concern, then, is not so much a description or prediction of the player's behaviour as it is the observation of *changes* in behaviour at different values of R.[5] R is assumed to be identical for Player I and Player II.

Taylor expresses the preference relations of the game as $Z \quad 0 \quad X \quad W \quad Y$ where Z and X are the criminal's payoffs from attacking an armed and an unarrmed victim, respectively, and W and Y are the armed and unarmed victim's payoffs from an attack (Taylor 1995: 271). This preference relation should immediately warn us that something

[5] For a thorough treatment of altering the nature of a game via external variables, see Heckathorn (1996).

is amiss; the relations $(X \quad W), (Z \quad W), (X \quad Y)$, and $(Z \quad Y)$ assume transferable utility, which except when explicitly dealing with a measurement of utility valued equally by both parties (for example, money) is an invalid assumption (Binmore 1992: 115). Luckily for us (or for Taylor!) the preferences which assume transferable utility are not compared with each other when payoffs are determined. The oversight *does* illustrate how easily errors can slip into otherwise-straightforward models.

More curious is $(0 \quad X)$, which suggests that a criminal will prefer not attacking an armed victim to being evenly matched with him, which is nonintuitive given that Taylor explictly notes that the criminal in the encounter has the "advantage of surprise" (Taylor 1995: 271). He uses this "advantage of surprise" in order to support $(X \quad W)$, though, so discarding that relation allows us to discard the imprecise "advantage of surprise", leaving a criminal preferring not attacking to being in a perflectly-matched fight, which, while not the only possibility, is at least a possibility. Having ensured that the surface imperfections in preference relations do not disturb the model, we can see where the model leads.

Results of the basic game Taylor correctly perceives that sufficiently large values of C make $[A, NG]$ and $[NA, NG]$ Nash equilibria (Taylor 1995: 273). In order to exclude these trivial cases, the range of C is restricted to values low enough to force mixed strategies.

In mixed strategies, Player I has a certain probability of being armed, and Player II has a certain probability of attacking; the Nash equilibrium is the point at which the respective probabilities balance out such that a player is indifferent about choosing one strategy over the other. Taylor's calculations of payoffs are technically accurate and merit no discussion here other than to note the payoff to Player I, which is the dependent variable which we wish to maximize in determining gun-control policy, that is, choosing a value for R. Thay payoff is $CY/(W \quad Y)$.

The derivative of that with respect to R is $Y/(W \quad Y)$ which is negative, leading to a startling observation: *any increase in the cost of obtaining a gun reduces the payoff for the victim.* The model is consistent with Polsby's conclusions. (In fact, the payoff function for Player I is discontinuous; when it no longer pays off for Player II to use a gun (*ie.*, when $Z \quad C$), Player I's payoff raises to 0 in the utopian outcome $[NG, NA]$. Taylor, assuming universal disarmament unrealistic, discards that outcome (Taylor 1995: 275); while not explicitly true to the model, this seems a reasonable restriction, especially in light of his

comment that the government cannot prevent someone from getting something they want, but can only make it more difficult (Taylor 1995: 272).)

Taylor's primary conclusion from the simple model is that any attempt at gun control will reduce social welfare (Taylor 1995: 275); "gun control tends to disarm victims and encourage criminal predation." Within the model this appears unequivocal. However, he concludes further that a move from the mixed-strategy Nash equilibrium to $[G, NA]$ would increase social welfare (Taylor 1995: 277). He begins his explanation, "Suppose that criminals could observe the armament decisions of individuals." Certainly, then, NG would be strongly dominated, and thus so would A, leaving $[G, NA]$; but that is an different game than the one analyzed, being no longer simultaneous-move nor of imperfect information.

It would also appear to require every man, woman and child to carry a gun and be able to use it. Since the minimum cost of having a gun is $C\ , C\ > 0$, this introduces a free-rider problem where one potential victim *not* being armed would diverge from $[G, NA]$, bringing us back into mixed-strategy equilibriums, but with a different and considerably more complex model. This *non sequitur* regarding $[G, NA]$ resides outside the model and is not supported by it; it appears to be the result of Taylor's political viewpoints regarding the concealed carrying of firearms (Taylor 1995: 277) leading to creative interpretations of the boundaries of the model.

(Note also that $[G, NA]$ may not be an equilibrium if we factor in the "advantage of surprise" that Taylor gives the attacker, which we—but not Taylor—discarded earlier; if evenly-matched armed encounters lead to a high enough payoff for the attacker, the equilibrium could just as easily be $[NG, A]$: where attackers have the upper hand in an attack, they will always attack, so victims—who lose either way—will avoid the cost of obtaining a gun and go unarmed. But this is as external to the model as is Taylor's universal civil armament outcome, $[G, NA]$, and should only serve to demonstrate the wildly erroneous conclusions which can result from the introduction of imprecise concepts—in this case, the advantage of surprise and observed armament decisions.)

The Complex Model In the second part of his article, Taylor attempts to account for some of the assumptions made in his basic model. In terms of method, this is practical; iteratively refining the model and observing the effects of the refinement on the outcome

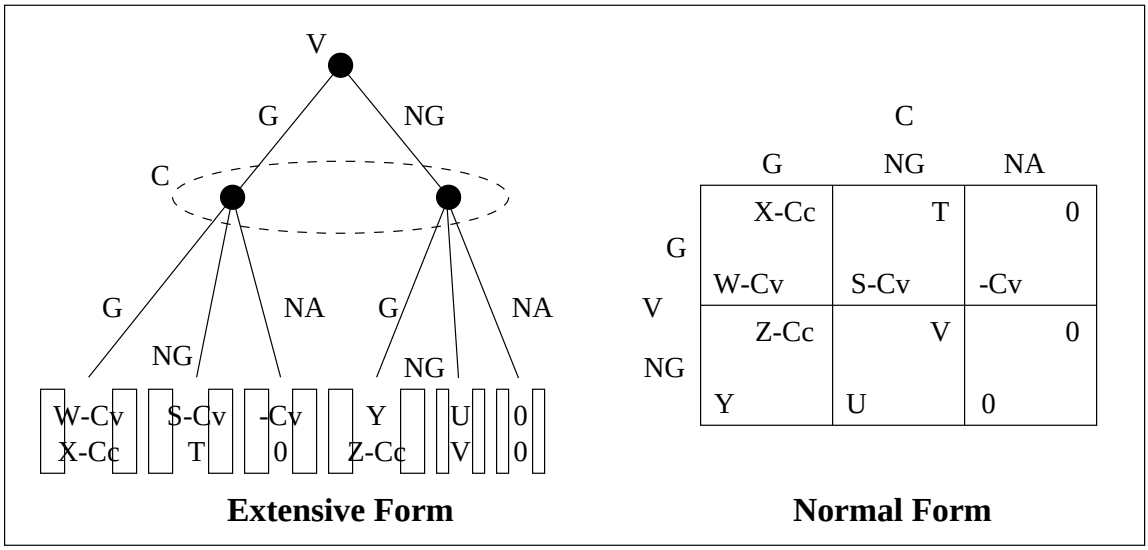

Figure 8: Taylor's complex gun-control game.

appears to be a particularly productive way of not only finding results but determining as soon as possible if the refinements provide useful results. Practically, however, an increase in complexity means an increase in the potential for imprecision and faulty assumptions, which when trying to both develop and prove a model *and* analyze its implications within one article are increasingly difficult to avoid.

The second model makes two important refinements on the basic one. First, Player II has a third option, NG, to attack without a gun; second, criminals can be targeted by gun-control measures, such that the cost of obtaining a gun for Player II (C_c) is higher than that for Player I (C_v). The game is illustrated in Figure 8.

Again, Taylor's preference relation suggests transferable utils, which is not the case, and again the relations which suggest that have no effect on the outcome. Player I's preferences are $0 \quad S \quad U \quad W \quad Y$, where S and U are an armed and unarmed victim's payoffs from an unarmed attack, and W and Y from an armed attack, respectively (Taylor 1995: 280). These seem unobjectionable. Player II's preferences are $Z \quad V \quad 0 \quad X \quad T$, where Z and V are the armed and unarmed criminal's payoffs for attacking an unarmed victim, and X and T those for attacking an armed victim (Taylor 1995: 278). Taylor again appears to order preferences based on the "advantage of surprise". If we discard that

along with the transferable utils, we still end up with questionable relations: $V \quad 0 \quad X$ suggests that an evenly-matched unarmed fight produces a higher payoff than an evenly-matched armed fight, for which Taylor gives no explanation. (Even if $V \quad 0 \quad X$ was chosen arbitrarily, explicitly saying so reduces the need for speculation that potentially important aspects of the model received insufficient attention.)

Results of the complex game For reasons as above, pure strategies are dropped from consideration (Taylor 1995: 280). This model ends up with two distinct mixed-strategy Nash equilibria; when R is low enough to make gun use efficient, Player II chooses between G and NA, and when R is high enough to make gun use inefficient, Player II chooses between NG and NA (Taylor 1995: 281). Further, the slope of the plot of Player I's payoff vs. R decreases (Taylor 1995: 282). In other words, by introducing NG and by using a higher R for Player II (criminal) than Player I (victim), we see the following developments: Player II will not attack with a gun when it means a sure loss to do so; Player II will not attack without a gun when it means a sure loss to do so; and Player I's welfare will not decrease as rapidly with increases in gun-control measures.

While Taylor seems to present these as important changes (Taylor 1995: 282), they appear a little tautological; furthermore, the payoff function with respect to R still has a discontinuity at the point at which Player II moves from G to NG as his "attack" strategy. Inexplicably, Taylor suggests that "modest gun control actually increases social welfare" (Taylor 1995: 283); the point at which it does so is the point at which all criminals are disarmed. It is not clear why this scenario, which was unattainable in the first model, is now attainable; since nothing has changed to allow it, logic would suggest that the condition mentioned by Taylor regarding government inefficacy still applies. This is to say, the point at which G is strongly dominated in the second model is the same as the point at which G is strongly dominated in the first, and that point will not be reached for the same reasons.

Contrary to his conclusions, the refined model appears to uphold the conclusion of the basic model—that since disarmament cannot be obtained, any gun control measures will reduce social welfare.

Evaluation With a well-constructed basic model, Taylor has discovered an aspect of gun-control policy which goes against common-sense perceptions; that alone makes his an important study and a practical application of game theory. His attempts at introducing a more realistic model indicated on their own that the conclusions from the basic model continue to stand as real-world complications find their way in; although his stated conclusions and those which the enhanced model presented diverged somewhat, the findings from the basic model still continued to stand. Through carelessness and lack of rigour in his second model, he effectively contradicts his initial results—luckily, the contradiction is weak enough to not shadow his confirmation of Polsby's theory.

That being said, the focus of this paper was gun-control *policy*—is it possible to go from this model to a workable policy? In this particular case, that doesn't appear to be possible. The model is too simple; as Taylor points out, something as simple as noting that criminals can attack other criminals (Taylor 1995: 287) complicates the question of who needs to own a gun and who will be doing the attacking enough to make this model inaccurate. He also notes that the assumption of risk-neutrality is probably inaccurate (Taylor 1995: 287); if the players in the game have different risk assessments, mixed strategies change, and the patterns observed above may no longer appear.

As well, the assumption made at the beginning of the paper—that "most arguments" about gun control claim that by moving closer to universal disarmament, social welfare increases—does not seem to be supported by the claims of gun-control advocates, who concentrate as much on accidental injury, suicide prevention, and gun culture as they do crime control (Spitzer 1998: 65; Carter 1997: 105). Paralleling that, the number of firearm deaths by suicide remains slightly *above* those from homicide (National Center for Injury Prevention and Control 2001: 2). Between the questionable premise of the study and the creative policy recommendations that resulted from it, it seems as though Taylor has spent much of his effort attacking a straw man—and when the premise of the experiment is invalid, the reliability of the outcomes come into question.

From this two further observations on the utility of game theory in policy development can be drawn. First, that by providing simplified models with well-defined preferences and measurable outcomes, game-theoretic models provide the tools to discover relationships between policy and performance which might otherwise escape (empirical) study, such as the declining-welfare effects measured here; but in doing so, such models are necessarily

unable to stand on their own as justification for policy. Lastly, that reaching a preliminary conclusion based on a simplified model may be considered a 'second-class' result in sociological or policy research in which real-world answers are the immediate goal. Fields in which that remains the case need to look to economics, statistical mathematics, and the hard sciences to see the utility of preliminary and abstract work; one can only hope that little work of value but too straightforward for publication falls through the cracks before that utility is noticed.

Discussion

From these five articles come two impressions: that game theory has much to offer to sociology, and is capable of being applied to research about things other than game theory itself, and that at the same time, it quickly becomes apparent that there is still much to be desired in the manner in which sociologists employ game-theoretical methodology in their work. While there certainly exist outstanding game-theoretical works, the problems of underapplication and misapplication of game-theoretical methods are real. Underapplication is almost certainly an effect of game theory's relative novelty in the sociological repertoire; as mentioned in the introduction to this paper, economists turn to game-theoretical methods ten times as often as do sociologists, and at least part of that must result from the economist's view that game theory is a valid, relatively unremarkable methodology. But misapplication is less straightforward; there are a definite set of areas in which sociologists' use of game theory appears, based on this admittedly small sample, to be lacking.

This is not to suggest that their work is without merit; while parts of it are hit-and-miss, much of the 'miss' material is still relevant at best and salvageable at worst, and the theoretical gains from the 'hits' are often of great importance—witness Kahana and Weiss's discovery of the moral hazards of *over*work in the profit-maximizing firm, Taylor's observations on the entirely *negative* impact of gun control, Ginkel and Smith's theory of frozen information, or Brown and Boswell's straightforward repair of split labour-market theory—but there is unquestionably work to be done. The problems that arose in the articles studied here tended to appear in the same areas: insufficiently rigourous models and calculations, a lack of empirical evidence, insufficient ties to the rest of the sociological literature, and accessibility to the game-theory neophyte.

Mathematical rigour The cause of the majority of the *surface* imperfections in the articles discussed above—and in one case of the unreliability of the authors' conclusions—is lack of attention to mathematical detail. Plainly, sociologists do not always come equipped to do the rigourous analysis necessary to generate a cohesive mathematical model. This is far from pervasive; Brown and Boswell, Kahuna and Weiss, and Ginkel and Smith encounter few mathematical hiccups in their work. But the effects that mathematical flaws introduce can be devastating; Schroeder and Rojas's imprecision leads to a model which is near entirely unreliable, and Taylor's model appeared to have survived his diversions

by a combination of luck and originality. The most common error was that of introducing factors which the model does not accommodate. It is as if the sociologist feels a need to introduce complications to overcome the purely economic viewpoint of his modeled actors. Ginkel and Smith most clearly addressed the problem of lenient analysis. By maintaining a table of variables, and by simplifying entirely through omission of detail, they were able to maintain a model free from contamination without sacrificing the cleanness, simplicity, and intuitiveness of their revolution game. Sociologists certainly have the skills to perform rigourous analyses; it should be a simple matter of taking the same approach to the elements of a game-theoretical model as one would to an ANOVA.

A related problem is that of *training* budding sociologists in game theory. While learning enough statistics to be able to perform correct analysis requires little more than high-school algebra, game theory requires a considerably more advanced grounding in math. At a minimum it requires an understanding of integral calculus, which amongst my undergraduate colleagues is rare. The problem of game-theoretical pedagogy is still being addressed in economics; similar pedagogical advice is beyond the scope of this paper. But a lack of integration with the general undergraduate sociological curriculum will, I expect, result in the situation where game theory finds itself constrained to those sociological research areas which resemble economics—organizational theory, industrial relations, labour markets, and work—simply because it will be those students who will have had attended enough courses taught in their institution's economics department to gain enough exposure to game theory to generate interest, even if instruction in it remains rudimentary. Integrating game theory into the sociological curriculum will be necessary if it is to spread from *homo economicus* to *homo ethicus* and beyond.

Empirical evidence Another element necessary in developing truly sociological applications of game theory is empirical support. Sociologists do not have the opportunity, afforded to mathematicians and economists, by which they can be satisfied with entirely abstract theory. Instead, there is an expectation that their theories will be based in real, observable experience. If economics is, as Binmore (1992) comments, the physics of the social sciences, then sociology is their biology; the 'spherical elephant in a vacuum' is insufficient. Some of the authors within grasped this; Brown and Boswell and Ginkel and Smith explicitly apply their models to historical events whose inputs, moves and outcomes

are known. But Kahana's otherwise impeccable study of absenteeism is left wanting for evidence that things are as they claim—*do* the Kibbutz and similar organizations present a moral hazard which requires sanctions to correct? Similarly, Schroeder and Rojas's risky-sex model is as presented far too fragile to survive without evidence that individuals looking for protected, no-commitment sex act strategically upon finding a partner similarly interested. Empirical support cannot be replaced with logical support, either, although the latter certainly has its own use; it is not enough to show that a model is consistent with itself and with common-sense expectations if the model, and those expectations, are nowhere to be found beyond the pages of the study in which they appear.

Embeddedness The models *themselves* must also stretch beyond the pages in which they are appear. The preponderance of economics-based work in the game-theoretical sociological literature hinted at above when discussing pedagogy also appears when observing the relationship between sociological research which employs game theory and the rest of the literature. Brown and Boswell demonstrate an ideal approach to embedding game-theoretical methods in the standard literature by applying game theory to a decidedly non-game-theoretical theory—and then, by expressing their results in terms of modifications to split labour market theory, they introduce their developments back into the literature such that those who rely on SLM—but not on game theory—can profit from them. Ensuring that links exist between the game-theoretical sociological literature and the conventional is a prerequisite to the widespread adoption of game-theory; that is, without sufficient embeddedness, game-theoretical approaches will remain a sort of hobby industry in sociology as they did in their early years in economics. At the same time, sufficient embeddedness will contribute to solving the problems of empirical evidence—by establishing connections to more conventional empirical work—and of accessibility, the availability of the work to those not versed in game theory.

Accessibility That game theory imposes certain mathematical requirements on its readers is obvious, but the seemingly unavoidable problem that creates must be addressed before game theory can take a place in standard sociological method—else many of those who do not already work with game theory may not be able to penetrate game-theoretical findings. We observed this obstacle most directly in Kahana and Weiss, above, in that by developing their absenteeism model directly from theorems of game theory—from what

might be called 'first principles'—they greatly reduced the ease with which their findings might be applied to conventional analyses of the firm. (Their trigger mechanism in particular is impractically complex; see Kahana and Weiss 1992: 582). It can also be an indirect obstacle; in Schroeder and Rojas's presentation, they intentionally omitted much of the mathematical content of their model, making it impossible to verify the manner in which they operationalized some of the elements they introduced while discussing their model.

On the other hand, both Brown and Boswell, and Ginkel and Smith acknowledged the accessibility problem and explicitly organized their presentation to benefit the sociologist who is a game theory layman. Ginkel and Smith's practice of relegating the derivations of their model's properties to an appendix is rare in game theory, but is not particularly innovative; after all, sociological researchers whose results rely on complex statistical analysis regularly relegate their math to endnotes and appendices, fully aware that the average reader will be more interested in their results than in their process. In doing so they also refrain from excluding those readers who do not have the ability to verify the calculations in the first place. Such a separation need not put accuracy of scholarship at risk; by relying on articles published previously (as in Brown and Boswell) or by including the technical parts outside the body of the article itself (as in Ginkel and Smith) it is still left up to the editors and referees of the publication in which it is to appear to ensure that standards of scholarship are maintained. It just no longer puts that requirement on the reader, while leaving open the option of verification.

The extreme case of ensuring that the substantive content of work that relies on game theory remains accessible to those without the mathematical prerequisite is, as was noted earlier, found in Brown and Boswell, who draw on Doug Heckathorn's collective-action model, thus enabling them to concentrate fully upon their strikebreakers. I referred to such an arrangement then as an *academic division of labour*, and that arrangement merits further discussion here.

Such an academic division of labour presents three outstanding advantages. Firstly, it *encourages the reuse of general formal models.* This practice enables a variety of problems in divergent areas of the social sciences to be addressed with the same model. This not only saves those who wish to apply the model the trouble of having to develop it first, but allows those formulating and developing the model to do so unencumbered by having to demonstrate practical results. At the same time, it encourages the game theorist to

develop a robust model; it is one thing to gloss over an assumption in the short exposition of a model being used once, but it is considerably less acceptable to do so when the intent of your work is to produce a general tool to be applied to many different problems by a variety of other authors. One interesting side effect of this practice is that it can lead to the detection of similarities within areas of research which at the surface are extremely dissimilar.

Second, it *encourages specialization*, such that a smaller number of social scientists need to be game theorists in order to take advantage of the insights which game-theoretical methods can produce. Both the requisite formal mathematical background and the ability to identify and construct abstract models are assets possessed by relatively few (and are assets which might draw one into disciplines outside the social sciences), while those who could benefit by having game-theoretical tools on which to draw in their practical research are greater in number. Leaving the development of sociological game theory to those whose research role is explicitly that thus produces positive benefits both on the robustness and innovativeness of the models produced and of the ease in which they can be integrated into research without having been the theorist who formulated them. Comparative advantage can work between individuals.

Lastly, it benefits the reader in that it *improves accessibility* by separating the technical formulation of the model from its application. Brown and Boswell provide a particularly effective example of this; the mathematical prerequisites for the reader of their article are reduced to the ability to interpret graphs, and a general understanding of equilibrium, costs and payoffs, and first, second, and third-order effects (all of which are addressed, incompletely but to an appropriate depth, by Brown and Boswell). Meanwhile, those interested in the model itself are referred to an appendix, or better yet, to Heckathorn's articles themselves. As anecdotal evidence of the propensity for papers based heavily upon game theory to present a barrier to the general reader, I recall an instance while preparing this paper in which both my advisor and a reference librarian expressed concern about the mathematical requirements necessary to understand an article, published in the *Journal of Mathematical Sociology*, that was under consideration for inclusion but was unavailable. Were such a division of labour as that which I recommend adopted there, the conclusions of the article—presumably, unusually mathematically intense—may have been more accessible to those who while sociological experts are mathematical laymen.

Two objections might be raised against such a division. First, it might be suggested that the circumstances arising in Brown and Boswell's study—in which a very well-documented historical event has been theorized upon, and that theory is expanded through game theory—might be so uncommon an approach as to make a poor platform on which to develop methodologies. As it happens, it is not necessary that the events be *historical*; only that there is empirical evidence to match not only results but the model itself. It is, as we saw in Schroeder and Rojas, possible to match results without matching the model, in which case having reached the correct results can be viewed as nothing more than coincidence at best or back-fill at worst; but if the model accurately represents a simplification, and not a distortion, of a real thing, and the outcome of the model matches the outcome of the real thing, then the conclusion cannot be said to be unreliable on the basis of methodology. But without such empirical evidence, the researcher is left with nothing to compare against: a game-theoretical model with no parallels in non-game-theoretical theories or in real events is a theory, not a methodology, and must be demonstrated correct outside of itself.[6]

The second objection is somewhat more serious: If such a division is adopted, would later work not be delicately balanced upon prior conclusions such that any flaws in the early work brings down all of the work based upon it, in the manner of a house of cards from which a card is removed from the base? The simple answer to this objection is "Yes, it would"—but with the qualification that it is no different than any *other* body of scientific knowledge in that respect. Social-science game theorists seem to have a tendency to prefer inventing their own wheels by generating models from first principles in each study, but elsewhere in the social sciences—and even more so in economics and the hard sciences—relying on the theoretical work of others is *de rigeur*. It is an expected risk of academic work. (Consider, for example, the effect upon the physicist if it were discovered that the theory explaining gravity were to be found to be inaccurate—yet a similar revolution has happened for nearly everything in physics *but* gravity, and the discipline survived.) Dividing academic labour in this manner produces no more risk than that which academia has already organized to avoid, while promising insight and productivity beyond that which could be accomplished by researchers working alone from first principles.

[6] Recalling, somewhat loosely, Godel's theorem of incompleteness?

Innovation In applying the recommendations suggested above, it is crucial to keep in mind that, while game theory has had the opportunity to develop and stabilize, it is still a very young *sociological* method, and that real innovation must be encouraged. Taylor's article provides a pointed example of the importance of this: despite starting with a shaky premise, tackling the development of the model himself, making assumptions and introducing factors that the model did not account for, limiting his embeddedness to a theory introduced in a mass-market magazine, and providing neither empirical support nor suggestions on how his model might be tested empirically, he managed to apply an innovative approach to find a possible relationship which traditional approaches were unable to detect.

As game theory slowly becomes established as sociological method, more innovative ways of applying it to the problems of the discipline will unquestionably come to surface. But first and foremost it must be kept in mind that game theory stands to offer the most to sociology if it is viewed as a *tool*. Despite being called a 'theory,' it is a means, not an end in itself—it is a theory of mathematics, not of society. Early adoption of game-theoretical methods will be by those most excited about its potential—but those same authors will be the ones with a tendency to write *about* game theory, rather than using it to advance the sociological literature *per se*. But if game theory is applied intelligently and productively to sociological problems—keeping in mind the recommendations above and the experiences results of the authors discussed within, as well as those not covered and the established body of game-theoretical work in other social sciences—then advances will undoubtedly be made which would not be reached through more conventional means. Practical game-theoretical methodology is off to a healthy start in sociology, and it is still early; with care, it has the potential to grow into a fresh, innovative, and productive basis for sociological research.

Appendix A

Strikebreaking or Solidarity in the Great Steel Strike of 1919: A Split Labor Market, Game-Theoretic, and QCA Analysis

Cliff Brown and Terry Boswell

Replace this with BROWN

Appendix B

A Game-Theoretic Model of HIV Transmission: Signaling and Coordination in a Game of Limited Information

Kirby Schroeder and Fabio Rojas

Replace this with SCHROEDER

Appendix C

So You Say You Want A Revolution: A Game Theoretic Explanation of Revolution in Repressive Regimes

John Ginkel and Alastair Smith

Replace this with GINKEL

Appendix D

Absenteeism: A Comparison of Incentives in Alternative Organizations

Nava Kahana and Avi Weiss

Replace this with KAHANA

Appendix E

A Game Theoretic Model of Gun Control

Robert Taylor

Replace this with TAYLOR

References

Aumann, R.J. (1986). "Game theory." *The New Palgrave Dictionary of Economics and the Law.* New York: Stockton Press, 1998. v.3: 460–482.

Axelrod, Robert (1986). "An evolutionary approach to norms." *American Political Science Review* 80, 1097–1111.

Binmore, Ken (1992). *Fun and Games.* Lexington, MA: D.C. Heath and Company.

Brown, Cliff and Boswell, Terry (1995). "Strikebreaking or Solidarity in the Great Steel Strike of 1919: A Split Labor Market, Game-Theoretic, and QCA Analysis." *American Journal of Sociology* 100(6), 1479–1519.

Carter, Gregg (1997). *The Gun Control Movement.* New York: Twayne Publishing.

Foster, William (1920). *The Great Steel Strike and Its Lessons.* New York: Huebsch.

Ginkel, John and Smith, Alastair (1999). "So You Say You Want A Revolution: A game theoretic explanation of revolution in repressive regimes." *Journal of Conflict Resolution* 43(3), 291–316.

Heckathorn, Douglas D. (1989). "Collective Action and the Second-Order Free-Rider Problem." *Rationality and Society* 1, 78–100.

Heckathorn, Douglas D. (1990a). "Collective Sanctions and Compliance Norms: A Formal Theory of Group-Meditated Social Control." *American Sociological Review* 55, 366–384.

Heckathorn, Douglas D. (1990b). "A Sequential Decision Model for Mixed Sanction Systems." *Policy Studies Working Group.* Working Paper no. 125. University of Missouri, Kansas City; Department of Sociology.

Heckathorn, Douglas D. (1996). "The Dynamics and Dilemmas of Collective Action." *American Sociological Review* 61, 250–277.

Kahana, Nava and Weiss, Avi (1992). "Absenteeism: A Comparison of Incentives in Alternative Organizations." *Journal of Comparative Economics* 16, 573–595.

Kuran, Timur (1995). *Private truths, public lies: The social consequences of preference falsification.* Cambridge, MA: Harvard University Press.

Morrow, James D. 1994. *Game Theory for Political Scientists.* Princeton, NJ: Princeton University Press.

National Center for Injury Prevention and Control (2001). "Firearm Injuries and Fatalities." *Fact Sheet Series.* Washington, DC: Center for Disease Control.

Polsby, D. (1993). "Equal protection." *Reason*, October 1993, 34–38.

Ragin, Charles C. (1987). *The Comparative Method.* Los Angeles: University of California Press.

Schroeder, Kirby D. and Rojas, Fabio (2000). "A Game Theoretic Model of HIV Transmission: Signaling and Coordination in a Game of Limited Information." Unpublished manuscript presented at the American Sociological Association Medical Sociology Session.

Spitzer, Robert J. (1998). *The Politics of Gun Control.* Chappaqua, NY: Seven Bridges Press.

Taylor, Robert (1995). "A Game Theoretic Model of Gun Control." *International Review of Law and Economics* 15, 269–288.