Critique of Schroeder and Rojas's "A Game Theoretic Model of HIV Transmission"

Rich Lafferty

1 Introduction

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 behavior—which on the surface seems irrational—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 behavior, they are less able to offer a meaningful *explanation* of it" (Schroeder and Rojas 2000: 3, italics in original). Like Brown, they begin with theories which they suspect are incomplete, going so far as to locate the aspects of sexual behavior 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 behavior 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.

2 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

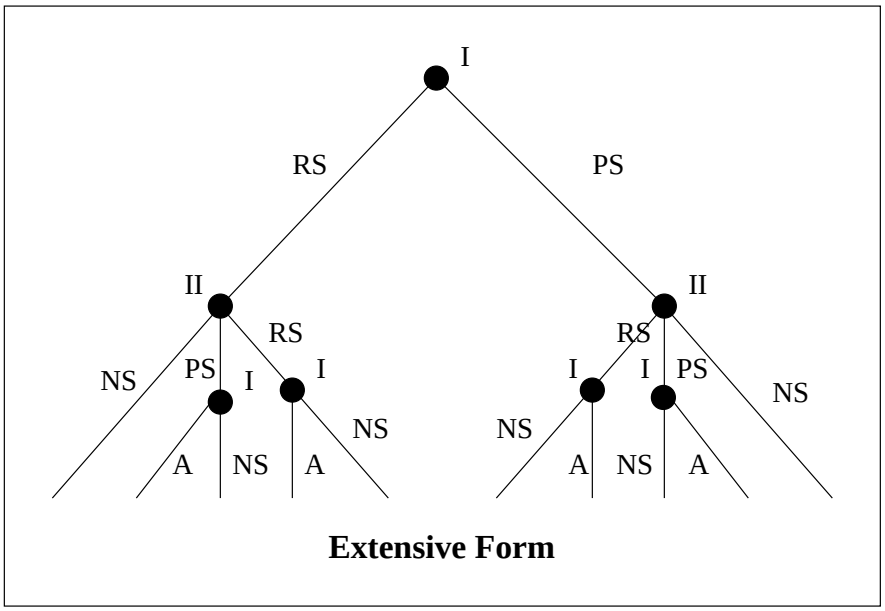

Figure 1: Schroeder and Rojas' (incomplete) risky sex game.

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 1. Nature's moves—which determine the HIV status of each player—and the outcomes of each play, are omitted from Figure 1 as well as in the figure provided by Schroeder and Rojas.

2.1 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, $RS > PS > NS$. For an HIV-negative player, $RS > PS > NS$ if the other player is HIV-negative, but $PS > NS > RS$ 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 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 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 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 1 omits Nature's moves (which determine the outcome of a play) and the payoffs (as compared to the moves) 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.

2.2 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 payoff $[NS, NS]$, 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' 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.

3 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 $h, h < 1$, so as to weight the tradeoff between immediate gratification and long-term gratification (Schroeder and Rojas 2000: 24). The game is repeated indefinitely (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 come into 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 (Schroeder and Rojas 2000: 16), altruistic condom use by HIV-positive partners with HIV-negative partners (Schroeder and Rojas 2000: 16), avoiding confrontation (25), extramarital sex and obtaining new partners (Schroeder and Rojas 2000: 25), 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.

4 Discussion

"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 (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 behavior 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 intoxicants 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 behavior, 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 behavior than we would be with those existing theories. 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.

5 References

Binmore, Ken (Schroeder and Rojas 2000: 1992). *Fun and Games.* Lexington, MA: D.C. Heath and Company.

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

Schroeder, Kirby D. and Fabio Rojas (Schroeder and Rojas 2000: 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.