Infidelity: The Prisoner's Dilemma

Paul David, Ph.D.

"To understand is to perceive patterns."

Isaiah Berlin
The Hedgehog and the Fox

Infidelity in monogamous relationships takes many forms, but it always involves partners pursing their own interests and breaking their promise to remain faithful. A game designed to examine the nature of self-interest in human interactions is the Prisoner's Dilemma. When applied to monogamous relationships, this game provides a mathematical model of the competing incentives involved in partners remaining faithful or cheating on one another. This article discusses the Prisoner's Dilemma, the game theory it is based on, and how this model informs us that self-interest is more about being faithful and conciliatory in the long run than pursuing our own selfish agendas in the short run. The focus of this discussion is on the logic rather than the mathematics of this model.

The Prisoner's Dilemma

The Prisoner's Dilemma was originally developed as a game by mathematicians to analyze and predict decision-making strategies when there are competing alternatives (Dawes, 1988). This game is called a dilemma because it poses a hypothetical scenario in which two imprisoned suspects are led to make an individual decision that is against their mutual interest. In its classic version, the police interrogate the two suspects separately. Despite the possibility of no conviction if both refuse to confess, each suspect is given the incentive of a lesser sentence to confess with the threat of a more severe sentence if one confesses but the other does not. As a result, it is in each suspect's self-interest to confess, but it is in their collective self-interest to hold out and not confess (Gottman, 2011).

Developed back in the 1960's, the Prisoner's Dilemma has morphed into hundreds of different games designed to examine the role that self-interest plays in decision-making (Poundstone, 1992). Although there are a myriad of scenarios, the basic structure of this game is the same. In any given round, there is a payoff of some kind, let's say six points, that can be evenly divided into three points for each of the two players, but what one player gets depends on whether the other player decides to split the points.

The challenge in this game is for each player to make a choice without knowing what the other one has decided. Again, deciding to cooperate means that both players receive three points. Deciding not to cooperate, or what game theorists refer to as "defection," results in a player obtaining at least one and possibly five points. However, if one player does decide to cooperate, that player runs the risk the other player will defect—collecting everything and leaving him or her with nothing. If both decide not to cooperate, they each receive one point. A summary of these different choices and combination of payoffs is delineated in the following matrix:

		Player B: Cooperation	Player B: Defection
Player A	Cooperation	(3,3)	(5,0)
	Defection	(0,5)	(1,1)

Given the uncertainty and payoff structure of this game, the logical solution is defection because a player who does not cooperate gets at least one point and possibly five points. In game theory, this particular choice is known as Nash’s equilibrium (named after John Nash who won the Nobel Prize in mathematics for his contributions to game theory). Nash’s equilibrium is the decision point where Player A’s best response is the same as Player B’s best response. As in these two-player games, the Nash equilibrium is not necessarily the most optimal outcome, but is the most advantageous decision the players can make when neither one is certain about what the other one will decide (Gottman, 2011). At the most abstract level, Nash’s equilibrium provides a solid mathematical basis for the logical choice between competing interests (Myerson, 1997).*

Because of its predictive capacity, this game theory model was initially applied to decision-making in economics and then eventually expanded to a wide range of other fields, ranging from political science to evolutionary biology (Camerer, 2003). Particularly when later research focused on the corresponding brain chemistry involved, it became

---

*For those interested in the mathematics of game theory, which are extraordinarily dense, there is a proof of Nash’s equilibrium for two-person games that is less elaborate than the more complex one that won the Nobel Prize. This proof goes as follows: Suppose that A and B are $m \times n$ matrices of real numbers and the strategy for Player A is a vector $p \in \mathbb{R}^m$ with

$$p_i \geq 0, \sum p_i = 1,$$

and the strategy for Player B is vector $q \in \mathbb{R}^n$ with

$$q_i \geq 0, \sum q_j = 1.$$

Nash’s equilibrium is a pair consisting of a strategy p for Player A and a strategy q for Player B such that for every strategy p' for A, $p' \cdot Aq \leq p \cdot Aq$, and for every strategy q' for B, $q' \cdot Bq' \leq p \cdot Bq$. The idea is that $p \cdot Aq$ is the expected outcome for Player A when A chooses strategy p and Player B chooses q. The first condition states that Player A cannot improve his or her outcome by unilaterally switching to some other strategy p'. Similarly, the second condition states that Player B cannot improve his or her expected outcome by unilaterally switching to some q'. Thus, given these conditions, Player A’s best response is the same as Player B’s best response (Coleman, 1999).

increasingly clear that the logic of decision-making is not as purely rational as this model had originally assumed. As explained in the next section, subsequent research on the Prisoner's Dilemma showed that decision making in humans is as much subcortical and limbic as it is cortical (Demasio, 1994; Lee, 2008; Guttman, Zeh, Pagnoni, Bems, & Kitts, 2002).

Fidelity in Intimate Relationships

As a game, the Prisoner's Dilemma is designed to explain how decision-making in human interactions is rule-governed (Camerer, 2003). When applied to fidelity in monogamous relationships, the numeric values in this game have no representational value other than to delineate the logical parameters of a couple's decision-making when the partners commitment to fidelity. In this respect, the Prisoner's Dilemma highlights the challenges often involved in being faithful in monogamous relationships. In these relationships, both partners attempt to insure a basic level of stability and predictability by agreeing to exclude other intimate relationships. In the context of game theory, they choose to mutually cooperate because it is the strategy that provides them the highest mutual payoff in the form of generating possible long-term attachment and security. However, like with the Prisoner's Dilemma, there is also the possibility of one partner taking advantage of the other partner's fidelity by selectively cheating because it has the highest individual payoff—a payoff that allows one partner to satisfy short-term interests (like sexual gratification) from a secondary partner while retaining the long-term benefits (like income security) from a primary partner.

Based on the reward structure of this game, where cheating has higher possible individual benefits, why do people decide to cooperate? Furthermore, in monogamous relationships, where partners still have the option of surreptitiously seeking out other opportunities, why do partners choose to remain faithful to one another? In other words, when self-interest is more likely to be rewarded in the short run, why do people try to cooperate and remain faithful in the long run?

The answer to this question is a complex one that involves multiple levels of influence. At the social level, one important factor is that humans are socialized to adhere to moral and ethical standards that value honesty and cooperation (Greenberg, Schmader, Arndt, & Landau, 2015). These standards are codified in civil and religious cannons that provide extensive social inducements and punishments to conform to these expectations. At the psychological level, another significant factor is that the affection and attachment humans experience with one another promote their desire to remain faithful (Fisher, 2004; Mikulincer, 2006). Because of the interdependence involved in these bonds, partners in intimate relationships place a very high premium on being able to trust one another. Consequently, when this trust is violated, it often leads to severe condemnation of the cheating partner and withdrawal of trust by the faithful partner (Gottman, 2011).

Another overarching influential factor is that evolution has selected for reciprocal altruism in humans. Unlike most other social species, humans have been sculpted by evolution to engage in mutual cooperation with both relatives and nonrelatives alike

(Axelrod & Hamilton, 1981; Trivers, 1971). As a result, part of the neural wiring in humans is designed to help them resist the temptation to selfishly accept but not reciprocate cooperation—particularly in regard to other family members (Rilling, Guttman, Zeh, Pagnoni, Bems, & Kitts, 2002).

A less apparent factor of why people remain faithful has to do with how reciprocal altruism is embedded in human brain chemistry. According to recent MRI studies of people playing the Prisoner's Dilemma, researchers found that when a subject cooperated, activity in the ventral striatum, the brain's reward center, would light up (Lee, 2008; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). In these studies, the ventral striatum was more sensitive to the total amount earned by both players, rather than to either player's individual accumulations.

What are the implications of these findings? First, there appears to be a basic altruistic component to the hard-wiring in our brains that derives more gratification from attending to the well-being of others rather than being solely concerned with our own welfare. Second, and even more importantly, along with the altruism there is a deep need for us to feel connected to our fellow human beings, and cooperation is the main survival mechanism that has evolved in the form of mirror neurons to produce this sense of connection (Dixit & Nalebuff, 2008; Pfaff, 2007).

In his MRI studies, researcher Mathew Lieberman (2013) found that our brains react to social pain and pleasure in much the same way as they do to physical pain and pleasure. He also found that our deepest pleasures are based on our ability to stay faithfully connected to the most important people in our lives. When particularly coordinated by the executive functions of the frontal cortex, this brain activity often leads us to restrain our short-term selfish impulses to preserve the longevity of our relationships (Kolb & Whishaw, 2011). These neural mechanisms lead to behavior that might seem inconsistent with our self-interest in the short run, but they are really about maintaining our well being in the long run.

Forgiveness in Relationships

The game theory studies on social dilemmas not only inform us about the nature of cooperation, they also highlight the central role that contrition and forgiveness can play in facilitating cooperation in relationships (Gottman & Silver, 2012). Here's why: when two subjects play one round of these games, they typically don't believe the other one will cooperate and, as a result, they predictably defect. The players are unfamiliar with each other and typically arrive at a Nash's equilibrium that fits their short-run perspective. However, when they play multiple rounds, the game gets more complicated in the sense that future cooperation typically depends on how the cheater reacts after he or she deceives the other player.

This reaction is an important variable in these games because if both players simply revert to a tit-for-tat pattern of retaliation (and they often do), it soon becomes clear to the players that their payoffs (1,1) will be minimal. In other words, if the cheating player is going to increase his or her chances of prevailing, he or she will have to cooperate even if the other player ceases to. Game theorists refer to this process as "contrition" and it eventually leads to another Nash's equilibrium based on trust and cooperation that results in consistently higher payoffs (3,3) for the players.

This process of contrition parallels what takes place in monogamous relationships that are attempting to recover from infidelity. As Gottman's (2012) research on repair of infidelity has shown, the unfaithful partner "must stick with the process and work to win back the other's trust, even if the partner doesn't respond at first" (p. 170). In their extensive study of people playing the Prisoner's Dilemma over hundreds of times, game theorists Robert Axelrod and Albert Chamman (1965) consistently found that contrition was a key variable in restoring trust and cooperation. This is not to say that the betrayed partners should blindly accept offers of future fidelity, but it does underscore the need for persistence and eventual acceptance of apologies as a critical step in recovering from infidelity.

Implications of Game Theory

Game theory models like the Prisoner's Dilemma provide a useful framework for understanding how individual decision-making in monogamous relationships can adhere to predictable patterns of self-interest. Originally designed by mathematicians to examine the logic of self-interest in decision-making, more recent studies applying game theory have demonstrated that self-interest is also influenced by subcortical and limbic brain chemistry that promotes honesty and reciprocal cooperation. The implications of these findings for monogamous relationships not only highlight how self-interest and faithfulness are amalgamated, they also point out the critical role that contrition and forgiveness can play when partners do cheat. As game theory reliably predicts, when contrition and forgiveness take place, faithfulness and cooperation tend to arise in future interactions—moving the relationship to a new Nash equilibrium based on trust. From this perspective, fidelity in monogamous relationships can be viewed as more about our contentment than our imprisonment.

References

Axelrod, R., & Chamman, A. (1965). *Prisoner's Dilemma: A study of conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

Alexrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*, 1390-1396.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Coleman, A. M. (1999). *Game theory & its applications*. New York, NY: Routledge.

Dawes, R. (1988). *Rational choice in an uncertain world*. Fort Worth, TX: Harcourt Brace Publishers.

Demasio, A. (1994). *Decartes' error: Emotion, reason, and the human brain*. New York, NY: Putnam.

Dixit, A., & Nalebuff, B. (2008). *The art of strategy: A game theorist's guide to success in business and life*. New York, NY: W.W. Norton.
Fisher, H. (2004). *Why we love*. New York, NY: Henry Holt & Co.
Gottman, J. M. (2011). *The science of trust*. New York, NY: W. W. Norton.
Gottman, J. M., & Silver, J. (2012). *What makes love last?* New York, NY: Simon & Schuster.
Greenberg, J., Schmader, T., Arndt, J., & Landau, M. (2015). *Social psychology: The science of everyday life*. New York, NY: Worth Publishers.
Mikulincer, M. (2006). Attachment, care-giving, and sex within romantic relationships: A behavior systems perspective. In M. Mikulincer & G. S. Goodman (Eds.), *Dynamics of romantic love* (pp. 23-44). New York, NY: Guilford Press.
Myerson, R. B. (1991). *Game theory: Analysis of conflict*. Boston, MA: Harvard University Press.
Kolb, B., & Whishaw, I. Q. (2011). *Brain and behavior* (3rd ed.). New York, NY: Worth Publishers.
Lee, D. (2008). Game theory and neural basis of social decision making. *Nature Neuroscience*, *11*(4): 404–409.
Lieberman, M. D. (2013). *Why our brains are wired to connect.* New York, NY: Crown Publishers.
Pfaff, D. (2007). *The neuroscience of fair play*. New York, NY: Dana Ptress.
Poundstone, W. (1992). *Prisoner's Dilemma*. New York, NY: Anchor Books.
Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Bems, G. S., & Kitts, G. D. (2002). A nueral basis for social cooperation. *Neuron*, *35* (2), 395-405.
Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). Opposing bold responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport*, *15*(16), 2539-2543.
Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35-57.