THE ABUSE OF PROBABILITY IN POLITICAL ANALYSIS: THE ROBINSON CRUSOE FALLACY

GEORGE TSEBELIS
University of California
Los Angeles

The decision to stay at home when you have no umbrella and rain is probable is an appropriate problem for decision theory. The decision to speed when you are in a hurry and the police might be patrolling is a game against a rational opponent. Treating the latter like a problem for decision theory is what I call the Robinson Crusoe fallacy. It is quite common and leads to incorrect conclusions. If the game has no pure strategy equilibrium, changes in the payoffs to a player affect not that player's strategy but the strategy of the opponent in equilibrium. For example, modifying the size of the penalty does not affect the frequency of crime commitment at equilibrium, but rather the frequency of law enforcement. I provide examples of this fallacy in regulation, international economic sanctions, and organization theory and argue that it stems from inappropriate use of the concept of probability.

> War is not an exercise of the will directed at inanimate matter. . . . In war, the will is directed at an animate object that reacts. It must be obvious that the intellectual codification used in the arts and sciences is inappropriate to such an activity.
>
> Clausewitz, *On War*

Robinson Crusoe was stranded on a desert island. He behaved as if he were the only person on the island until one day he discovered a footprint on the sand. From that moment he realized that he was not playing against nature any more. He was facing rational opponents, so he started fortifying his house to resist attacks. Clausewitz' statement indicates that mistaking a rational opponent for nature is an "inappropriate activity." I will call it a fallacy—the Robinson Crusoe fallacy.

My purpose is to generalize Clausewitz' statement for games that are not zero sum and to show that the Robinson Crusoe fallacy can lead to important mistakes. I will show that the expected utility calculations typically used in decision theory are inappropriate when probabilities are not exogenous but are part of the (equilibrium) strategy of a rational opponent. Mistaking the equilibrium mixed strategy of the opponent for a probability distribution is not an innocuous simplification. It leads to such important mistakes as the belief that modification of the payoffs of one player will lead to modification of that player's behavior. In reality, it leads to the modification of the (equilibrium) strategy of the opponent. I will provide examples of the Robinson Crusoe fallacy from the literature concerning crime, regulation, executive-legislative relations, international economic sanctions, and organization theory.

I first present two problems to which decision theory is traditionally applied: the decision to stay home or go out when rain is probable and the decision to violate or not violate traffic laws in the probable presence of the police. The former scenario correctly belongs to the domain of decision theory, while the latter is an example of the Robinson Crusoe fallacy and should be studied in a game-theoretic framework. Then I analyze the traffic violation example as a game-theoretic problem, demonstrating that the conclusions of decision theory are inappropriate, and those of game theory are counterintuitive. Finally I present examples of the Robinson Crusoe fallacy and put forward the argument that the concept of probability very often represents mixed strategies, with the result that the use of decision theory is inappropriate and may lead to wrong conclusions.

Rain and Crime: Decision Theory and the Robinson Crusoe Fallacy

Consider the following problem: rain is probable and you have no umbrella. How will you proceed in making the decision to stay at home or not?

According to Savage's (1954) terminology, there are two possible states of the world: rain and no rain. There are also two possible acts: to stay at home or to go out. Therefore, there are four possible outcomes: (a) you go out and get wet, (b) you go out and stay dry, (c) you stay home while it rains, and (d) you stay home while it does not rain.

Table 1 represents the utilities of these four possible outcomes. It is reasonable to assume that you prefer to go out if it does not rain. It is also reasonable to assume that you probably prefer to stay in if it rains. Formally, these assumptions can be stated, $b > d$ (A1) and $c > a$ (A2). In addition, it is very likely that there is a preference for rain if you stay in, justifying your choice ($c > d$). This assumption, however, is not necessary for what follows.

Table 1. Payoffs in the Rain Decision Problem

Decision	Rain	No rain
Go out	a	b
Stay in	c	d

Note: There are two states of the world (rain, not rain) and two acts (go out, stay in). Payoffs: $c > a$ and $b > d$.

In addition to assumptions A1 and A2, you need the probability of rain in order to make up your mind. Suppose that the weather report this morning gave you a reliable estimate p of this probability. Elementary decision theory dictates how to make your choice: you compare the expected utilities (EU) from each act and choose the act with the higher expected utility. In formal terms, you should go out if

$$EUout - EUin = [a \times p + b \times (1-p)] - [c \times p + d \times (1-p)] > 0.$$

In the opposite case, you should stay in.

In a series of similar situations, the probability that you will go out increases when the probability of rain decreases. Moreover, the probability that you will go out varies with your utilities: it increases with a and b, and decreases with c and d. Therefore, if you do not mind getting wet (a is high, since a is negative), the probability that you will go out increases. Similarly, if you intensely dislike staying in while it does not rain (low d, since d is negative), the probability that you will go out increases.

Consider another similar situation: you are driving your car and you are in a hurry. However, you do not know whether you should take the risk and exceed the speed limit. There are two states of the world: either the police are nearby or they are not. There are two acts to choose from: either to violate the speed limit or to abide by the law. Again, there

Table 2. Payoffs in the Crime Decision Problem

Decision	Police	No police
Speed	a_1	b_1
Not speed	c_1	d_1

Note: There are two states of the world (police, no police) and two acts (speed, not speed). Payoffs: $c_1 > a_1$ and $b_1 > d_1$.

are four possible outcomes: (a) you can get a ticket for speeding, (b) you can get to work on time without any incident, (c) you can arrive late and avoid a ticket, and finally (d) you can arrive late through there were no policemen on the streets.

Table 2 represents the utilities of these four possible outcomes. It is reasonable to assume that you prefer to speed if there are no police around. It is also reasonable to assume that you prefer not to speed if the police are out in force. Formally, these assumptions can be stated, $b_1 > d_1$ (A1′) and $c_1 > a_1$ (A2′). Again, even though it is very likely that there is a preference regarding the state of the world when speeding or observing the law, such assumptions are not necessary.

Aside from assumptions A1′ and A2′, the probability that the police are lurking nearby is necessary to know in order to arrive at a decision. Suppose an estimate p of this probability is obtained from experience. Elementary decision theory dictates how the choice is made. The expected utilities from each act are compared and the act with the higher expected utility is chosen. In formal terms, you should speed if

$$EU_{speed} - EU_{not} = [a_1 \times p + b_1 \times (1 - p)] - [c_1 \times p + d_1 \times (1 - p)] > 0.$$

In the opposite case, you should observe the speed limit.

In a series of similar situations, the probability of violating the law increases when the probability of lurking policemen decreases. Moreover, the probability of speeding varies with your utilities: it increases with a_1 and d_1 and decreases with b_1 and c_1. In particular, the size of the penalty for speeding will influence the decision to speed. A higher fine for speeding will decrease the frequency of violating the law.

The use of decision theory in describing a problem that "obviously" belongs to the domain of game theory, since nature and the police (a rational agent) are quite different in their behavior, is inappropriate. In other words, I have committed the Robinson Crusoe fallacy. However, this is a crude and oversimplified—yet faithful—account of the logic of what is known in the literature as the economic approach to crime, or deterrence theory. For authoritative accounts of these theories, refer to Becker 1968, Stigler 1970, Ehrlich 1973, and Ehrlich and Brower 1987, in which the conclusions of my simplified example—in particular the deterrent effect of the size of the penalty on crime—are stated and proven.

Why is a decision-theoretic approach appropriate for the problem of rain and inappropriate for the problem of crime? The obvious reason is that nature can reasonably be approximated by a probability distribution, while a rational player cannot. This answer pushes the problem a bit further: can we approximate a rational player's expected actions by a probability distribution?

Axiom P2 of Savage's account of expected utility theory provides the answer. This axiom expresses formally "the sure thing principle,"[1] and assumes that states of the world are independent of the acts chosen by the decision maker.

Take Newcomb's paradox (Eells 1982) as an example to demonstrate the meaning of independence between acts and states of the world. Suppose there are two boxes, one open and one closed. There is a thousand dollars in the open box, while the closed box contains either a million dollars or nothing. You are presented with the option of either taking the contents of both boxes, or only taking the contents of the closed box. There are two states of the world (the closed box has either a million dollars or is empty) and two acts from which to choose (either take both boxes or only the closed one).

There is a dominant strategy in this problem: take both boxes, because in every possible situation this choice offers a thousand dollars more. Formally, if we assign probability p to the situation where the closed box contains a million dollars, for any value of p in the [0,1] interval

$$EU_{both} - EU_{closed} = p \times 1000 + (1 - p) \times 1000 = 1000.$$

Now consider Newcomb's complication: assume you are informed that a demon decides the contents of the closed box; if it expects you to pick up only the closed box it puts a million dollars inside. If it expects you to be greedy and pick up both boxes, it places nothing in the box. Informed that the experiment has been performed a million times and that people *always* left either with a million dollars or with a thousand dollars, what will you do?

One possible line of reasoning is that whatever the contents of the closed box, those contents already have been decided by the demon and therefore your decision will have no effect. Consequently, you are better off if you choose both boxes. This reasoning applies the "sure thing principle" (Savage's axiom P2), a principle that chooses to ignore information that may be relevant. The second possible line of reasoning is that it is almost certain (because of the regularity of the results of past experiments) that the demon is always right in its anticipation: you will finish with a thousand dollars if you choose the dominant strategy and a million dollars if the dominated strategy is chosen. This reasoning ignores the "sure thing principle" because it assumes that states of the world are dependent on the acts chosen.

Put the insights generated by this (unrealistic) thought experiment in more plausible settings and consider the following questions:

1. Does the probability of rain change if people go out or stay in?
2. Does the probability of police enforcement change if people abide by the law or violate it?

The answer to the first question is *no*, while the second is *yes*. Indeed, the police will enforce the law if it is violated, preferring to relax if there is no reason to be present.[2] In other words, the conditional probability that the police will enforce the law—given that it is violated—is different from the unconditional probability of law enforcement. Thus, we cannot replace the action of the police by a probability distribution *independent* of the acts chosen by the people. Robinson Crusoe is not alone, doing the best he can against nature. We have to introduce Friday into the picture. In other words, not only do the police have to be modeled explicitly, but the probabilities of their actions have to be *derived* in the same manner as the probabilities of the public: through rational calculations.

A Game-Theoretic Analysis of Crime

Table 3 presents the payoffs of the police considered as a rational player. In order to find the ranking of these payoffs, I will proceed by way of assumption, as I did for the public. Assume that the police prefer to enforce the law when it is violated and that due to enforcement costs, the police prefer not to enforce the law if it is not violated. Formally, these assumptions can be stated, $a_2 > b_2$ (A3′) and $d_2 > c_2$ (A4′).

If the police enforce the law, the public will stop violating it (A2′), if the public stops violating the law, the police will stop enforcing it (A3′), if the police stop enforcing the law, the public will violate it (A1′), if the public violates the law the police will enforce it (A4′), if the police enforce the law, then the public will stop violating it (A2′), and so on. In other words, the conditional probability that the police will enforce the law (given its violation) is equal to one, while the conditional probability that they will enforce the law (assuming it is not violated) is zero. Similar statements concerning conditional probabilities can also be made for the public. These statements indicate the fundamental criticism mentioned in the previous section: that actors maximize against each other and not against nature, so their acts are not independent.

In this game between the police and the public, no matter which combination of strategies results from the choices of the two players, one of them will have the incentive to modify the choice. In game-theoretic terms, the game presented in Table 3 has no pure strategy (Nash) equilibrium since the two players have no pure strategies that are optimal responses to each other. Therefore, the only equilibrium strategies that exist in this game are mixed strategies that are probability distributions over the set of pure strategies (Ordeshook 1986, chap. 3). In other words, each player will use a combination of the pure strategies so that when both apply a specific pair of strategies, neither will have an incentive to deviate from his or her mixture. Each player's strategy will be the optimal response to the other player, therefore, both players will retain these (equilibrium) strategies.

In terms of probabilities, the fundamental difference between the decision-theoretic and the game-theoretic approach is that probabilities are not given exogenously (as in the example of rain), but come as the derivative of rational calculations on the part of the actors: acts are interdependent and therefore the conditional probabilities of acts are different from the marginal probabilities.

Table 3. Payoffs in the Police-Public Game

Decision	Enforce	Not enforce
Speed	$a_1\, a_2$	$b_1\, b_2$
Not speed	$c_1\, c_2$	$d_1\, d_2$

Note: Each player has two strategies. Payoffs of the public: $c_1 > a_1$ and $b_1 > d_1$. Payoffs of the police: $a_2 > b_2$ and $d_2 > c_2$.

In order to compute equilibrium strategies we must (1) assign a probability p to the public that chooses to violate the law (and $1 - p$ to those who abide by it) and a probability q to the police who enforce the law (and $1 - q$ to those who do not enforce it) and (2) find a pair (p^*, q^*) of p and q with the quality that the best answer for the public (provided the police follow the mixed strategy, specified by q^*) is to mix its pure strategies using p^* and $1 - p^*$ as weights; while the best answer for the police (provided the public follows the mixed strategy, specified by p^*) is to mix their pure strategies using q^* and $1 - q^*$ as weights. The calculation of p^* and q^* gives (Luce and Raiffa 1957)

$$p^* = (d_2 - c_2)/(a_2 - b_2 + d_2 - c_2) \quad (1)$$

$$q^* = (b_1 - d_1)/(b_1 - d_1 + c_1 - a_1) \quad (2)$$

We can verify that the specified probabilities p^* and q^* are in the (0,1) open interval and are therefore acceptable as solutions because of assumptions 1′–4′. Moreover, since the game has no pure strategy equilibria, the mixed strategies specified by the probabilities p^* and q^* are the *unique* equilibrium strategies of the game.

This last proposition is very important.

It must be singled out and stated formally.

THEOREM 1. Under assumptions 1′–4′ the only equilibrium in the police-public game is in mixed strategies as specified by equations 1 and 2.

This specification of the problem enables us to study the impact of different policy measures on crime. For example, what happens if the legislator influenced by the arguments of economic analysis increases the penalty for a crime? The size of the penalty is represented in the model by the payoff a_1, which the public receive when they violate the law and the police enforce it. Examination of equations 1 and 2 indicates the following important and counterintuitive outcomes.

THEOREM 2. An increase in the penalty leaves the frequency of violation of the law at equilibrium (p^*) *unchanged.*

THEOREM 3. An increase in the penalty *decreases* the frequency that the police enforce the law at equilibrium (q^*).

The proof of these two theorems is straightforward: inspection of equation 1 indicates that p^* does not depend on a_1, while equation 2 indicates a monotonic relationship between a_1 and q^*. What is needed, however, is to stress the reason for the resulting discrepancy between conventional wisdom and decision theory on the one hand and game theory on the other: in game theory both players maximize simultaneously while playing against each other. Neither of them can be assimilated to a probability distribution over states of the world. While it is very plausible that when penalties increase, the public (in the short run) will reduce violations of the law, it is also plausible that once the police realize this change in criminal behavior they will modify their own strategy; that is, they will try to reduce the frequency of law enforcement. The public will then modify again, and then again, and the new equilibrium will be the one described by equations 1 and 2 in which the increase in the penalty has no impact on criminal behavior.

Other policy prescriptions stemming from a sociological approach to the problem of crime suggest the development of welfare policies so that crime will cease to be an attractive career (Tittle 1983). In terms of the game-theoretic model, such policies would represent an increase of c_1 or d_1 or both—the payoffs for not violating the law whether the police enforce it or not. Note that the logic of these policy prescriptions is very similar to the logic of the economic approach to crime: they reward virtuous behavior instead of punishing deviance. In both cases, however, the expected utility of crime relative to compliance with the law is expected to decrease.

Examination of equations 1 and 2 indicate the following consequences of increasing welfare policies:

THEOREM 4. An increase of either c_1 or d_1 (welfare measures), leaves the frequency of violation of the law at equilibrium (p^*) *unchanged.*

THEOREM 5. An increase of either c_1 or d_1 (welfare measures) *decreases* the frequency that the police will enforce the law at equilibrium (q^*).[3]

The intuition behind these theorems concerning the sociological approach to crime is exactly the same as before: although in the short run such measures will have the desired impact, this will lead the police to modify their strategy and enforce the law less frequently. The public, in turn, will increase the frequency with which they violate the law. Thus, the process will finally equilibrate at the strategies described by equations 1 and 2.

The conclusions of the game-theoretic analysis reported in theorems 1–5 are strongly counterintuitive. It is therefore only reasonable to question the logical and empirical foundations of the validity of the game-theoretic approach. I will begin with the logical evidence. There are several specifications that lead to the outcome of equations 1 and 2.

Rational players with complete information. The equilibrium of the public-police game is unique. Therefore, it can be shown to have all the desirable game-theoretic properties of stability. In particular, this Nash equilibrium can be shown to be perfect (Selten 1975), proper (Myerson 1978), sequential (Kreps and Wilson 1982), and stable (Kohlberg and Mertens 1986) (see Tsebelis 1987a). Therefore, there is good reason to believe that rational actors will choose the equilibrium strategies described by equations 1 and 2.

Rational players with incomplete information. Suppose now that both players are rational but that each one of them knows only his or her own payoffs and does not know the payoffs of the opponent. If each player knows that there is a random element in the opponent's payoffs, the frequency with which this player will choose each one of the pure strategies tends to the equilibrium described by equations 1 and 2 as uncertainty tends to zero.[4] A natural interpretation of this approach would be that different police agents with different payoffs interact with different members of the public with different payoffs and that the frequency of choice of different strategies tends to the outcome described by equations 1 and 2 as these differences tend to zero.

Evolutionary approach. The same equilibrium can be supported under much weaker assumptions of adaptive behavior by both players through an evolutionary approach (Tsebelis 1987a). In this case individual actors are supposed to fulfill not the extraordinary requirements of rational choice theory but those of myopic adaptive behavior, leading them to choose at any time the strategy that pays better. So if the interaction of the public with the police is modeled along the time dimension, if after each time period each member of the public compares the results of abiding with the law with violating it and selects the most profitable behavior and if the police do the same (compare whether it is better for them to enforce or not to enforce the law), the equilibrium of this game is given by equations 1 and 2.

Mixed approach. It can be shown also that in the more realistic situation in which one of the players is fully rational (the police in this case, since it has a central organization and discipline) while the other is myopic (the unorganized public), the same strategies prevail at equilibrium.[5] The reason is that if the police choose any frequency q higher than q^*, the public will stop violating the law altogether, and the police will want to reduce the frequency of law enforcement. And if the police choose any frequency lower than q^*, the public will violate the law always, which will make the police increase q. So any frequency different from q^* cannot be an equilibrium. The relevant technical concept in this case is the Stackelberg equilibrium. But the Stackelberg equilibrium has a wider use: it is applicable when the two players move sequentially instead of simultaneously. So equations 1 and 2 indicate the equilibrium strategies of the opponents even in the case of a sequential game. A sequence of moves is not the natural interpretation of the police-public game, but it is the most appropriate assumption in other similar games, such as economic sanctions between nations, where the target country moves first and decides whether to violate some rule, norm, or economic interest of the sender country, and the latter decides whether to or not to punish. In such sequential games, the equilibrium is described by equations 1 and 2.

The last two specifications indicate that the assumption that the public is a unified rational actor capable of strategic behavior is not required to support the equilibrium of equations 1 and 2. Indeed, in the last two cases, the public—or more precisely *each member* of the public—is assumed to maximize its own payoffs without any strategic considerations. The fact that the game leads to the same outcome with such diverse specifications as perfect rationality, adaptive behavior, and complete or incomplete information is an indication of the robustness of the equilibrium (equations 1 and 2).

The empirical evidence concerning the absence of impact of penalty on crime is indirect. Although there is an abundance of empirical studies on crime and the impact of the frequency of law enforcement has been established, the impact of the size of the penalty has not.[6] Some empirical studies find a negative correlation between frequency of crime and size of penalty while others find a positive correlation. The explanation for the latter may be that when the frequency of crime increases, the legal system is likely to become more strict. Nagin, reviewing the empirical literature on the deterrent effect of penalties on crime for the National Academy of Sciences, concludes, "Yet, despite the intensity of the research effort, the empirical evidence is still not sufficient for providing a rigorous confirmation of the existence of a deterrent effect. Perhaps more important, the evidence is woefully inadequate for providing a good estimate of the magnitude of whatever effect exists. . . . Any unequivocal policy conclusion is simply not supported by valid evidence" (1978, 135–36). This empirically based statement is quite similar to theorem 2, which was derived from the game-theoretic approach. So the absence of clear-cut evidence in favor of this particular aspect of the economic theory of crime is an indirect empirical indicator of the validity of the game-theoretic approach.

The logical and empirical evidence in favor of the game-theoretic approach indicates that both dominant approaches to crime prevention—the sociological and the economic—are partial and short-run in nature. They are both based on the assumption that modifying the incentive structure of a rational agent (modifying the payoffs) will affect that agent's behavior. Reasonable and plausible as this assumption may seem, it was shown not to withstand game-theoretic scrutiny.

To recapitulate, I considered the problem of traffic law violations not as a problem in decision theory but as a game between the police and the public. The superiority of the second approach over the first arose from the fact that the probability of law enforcement was not given exogenously (as in the rain example) but derived as a conclusion from the rationality assumption and from the payoffs of the players. Modifying the penalty had no impact on the frequency of criminal activity, as is commonly assumed. Instead, the frequency of law enforcement is affected.

How frequently can common sense conclusions and widely held beliefs—such as the deterrent effect of penalties—be shown false under closer game-theoretical scrutiny? How common is the Robinson Crusoe fallacy? I argue that such situations are very frequent.

The Abuse of Probability and Decision Theory

Do the counterintuitive results of the previous section require unusually strong or unrealistic assumptions? Are they an artifact of the 2×2 game?

I have already demonstrated that the equilibrium solution holds even if we relax the requirement of strong rationality and replace it with one-sided rationality, mutual myopic behavior, or conditions of incomplete information. However, the absence of an impact of penalty on crime may be due to the restricted strategy spaces of the two players. What happens if each player has more than two strategies?

Let us consider more serious crimes and assume that the criminals have the option to commit more than one crime and that the police have the option of dividing their law enforcement activities among the different crime activities. Construct a payoff matrix for this game similar to Table 3, in which each player has $n + 1$ options: the criminals have n different crimes and obedience to the law (which I will call the criminals' zero option); and the police have enforcement of the law for each one of these crimes and no enforcement (which I will call the police's zero option). This situation is more realistic than the simple 2×2 original game. Note that each player now can choose not only in a binary way (zero option or not) but has a choice between different ways of spending time as well as all the combinations among them (mixed strategies).

Studying such a game will provide information about the consequences of modifying the penalty of one type of criminal activity. However, I will not construct and solve such a game. Instead, I will indicate the reasons why such a modification of one or all the penalties has, in this general case, consequences similar to those of the 2×2 case. Such a modification has no impact upon the behavior of the criminals, but it influences the behavior of the police.

To calculate the Nash equilibria of this game one has to verify whether pure strategy equilibria exist. If the game has a pure strategy equilibrium, the players will choose their pure equilibrium strategies. It is unlikely that this $(n + 1) \times (n + 1)$ two-person game has one pure strategy equilibrium, because in this case this equilibrium would still be an equilibrium in the 2×2 game in which the available options for each player would be the particular crime and the zero option. I have demonstrated, however, that a pure strategy equilibrium in the 2×2 game is very unlikely.

Nonetheless, it is possible that the $(n + 1) \times (n + 1)$ game has several pure strategy equilibria, which are combinations of mutually optimal strategies. In this case, we would observe the criminals and the police coordinating their activities in order to achieve one of two things: (1) when the criminals violate one particular law, the police are always there to enforce it, or (2) the police are never there when a crime is committed. Again, the arguments of the second section indicate the unlikelihood of such a situation.

Thus, the only remaining case is that the $(n + 1) \times (n + 1)$ game has no pure strategy equilibria. Two theorems can be demonstrated in this case.

THEOREM 6. If a two-player game has no pure strategy equilibria, it has a unique mixed strategy equilibrium.

THEOREM 7. In a two-player game with no pure strategy equilibria, modification of the payoffs of one player will lead that player either to change the pure strategies that he or she mixed or to leave the mixed equilibrium strategy unchanged and modify the equilibrium strategy of the opponent.

The existence of such an equilibrium has been proven by Nash (1951). The uniqueness and theorem 7 follow from the method in which this mixed strategy equilibrium is computed. Each player must choose a strategy mixure that will make the opponent indifferent as to his or her own strategies. Therefore, each player has to choose the probabilities of his or her own mixed strategy in such a way that the expected utilities of the other player's strategies will be equal. These calculations lead to a linear system of n equations with n unknowns (the probabilities that each player will make use of the different strategies) in which the coefficients are the payoffs of the *other* player. This remark proves theorem 7, while the uniqueness outcome of theorem 6 follows from the linearity of the system.

Now that the mathematical robustness of the outcome has been established, I focus on another, more interesting question. What is the frequency of games like the general $(n + 1) \times (n + 1)$ game I have just described? I will demonstrate that cases of the Robinson Crusoe fallacy—that is, situations in which decision theory is (mis)used in the place of game theory and leads to incorrect results—are very frequent, by providing some examples of such situations. Finally, I will give a set of sufficient conditions for such a misuse to occur.

Consider the case of regulated industries. Their situation is similar to the public of our 2×2 game while the regulating agency plays the role of the police in our 2×2 game. Industries prefer to abide by the law if they are going to be investigated and prefer to violate it when they know they will not be caught.[7] Agencies prefer to enforce the law when it is violated and to relax when industries conform to regulations. Again, most of the time there is no pure strategy equilibrium. Therefore, the unique existing equilibrium is in mixed strategies with the properties we have discussed: changes in fines or penalties do not affect the behavior of firms at equilibrium. This result is in sharp contrast with the economic literature on regulation.[8]

Consider the problem of international economic sanctions and the big dispute concerning their effectiveness. There seems to be a consensus among scholars that sanctions do not work.[9] Despite scholarly opinion, economic sanctions are applied with increased frequency over time.[10] If economic sanctions work, why do scholars suggest they do not? And if they are not effective, why do policymakers insist upon imposing them? The puzzle can be solved in terms of our simple model. There are two possible cases: either the penalty imposes on the target country costs more serious than the benefits of deviant behavior, or it does not. In the first case, there is no pure strategy Nash equilibrium, and the model developed in the second part applies, making the behavior of the target country independent of the size of the penalty. In the second case, there is a pure strategy Nash equilibrium, and deviant behavior should be observed all the time and no penalty imposed.[11]

Consider problems of hierarchies. Can the higher level be sure that lower levels will conform to given instructions? Should it increase rewards for observation of the internal rules of the organization? Should it punish deviant cases in an exemplary way? Should it increase monitoring frequency? In other words, which policy works better—the carrot or the stick—and how often should it be applied? Again, if the payoffs of the two players are such that there is no pure strategy Nash equilibrium, the size of the rewards or the penalties will have no impact on the behavior of subordinates at equilibrium. However, higher rewards and penalties will reduce the required monitoring frequency. Therefore, if the higher level of hierarchies can design the rules of the game (rewards and punishments), they will try to increase both not in order to reduce deviant behavior but to reduce the required monitoring frequency. The question of endogenous changes of the rules can be dealt with in terms of a two-stage game: in the first stage the higher level sets the rules (the payoffs), and in the second stage the actual game is played with simultaneous moves by higher- and lower-level agents. Given that the equilibrium in the second stage can be calculated exactly as in the game between the police and the public, one can work the solution of the game backwards, substitute the equilibrium payoffs for the actual games, and then choose between different games according to some utility function. For example, the higher levels of a hierarchy may want to consider a trade-off between monitoring frequency and levels of anxiety generated by high penalties. The higher levels of a hierarchy are thus able to maximize their utility no matter what the specific utility function is and what trade-offs the utility function establishes.

Finally, consider the problem of legislative oversight. It has been argued that Congress prefers the "fire alarm" system, where citizens or groups monitor the executive branch, as opposed to the "police patrol" system, which requires more active monitoring, because the former system is more cost-efficient (McCubbins and Schwartz 1984). However, empirical findings indicate that fire alarms do not constitute the majority of cases of legislative oversight (Aberbach 1987). The police-public game can help investigate this discrepancy. Indeed, the model can include three different strategies of the law enforcing agency: fire alarms, police patrols, and no action. Explicit use of the relevant payoffs will indicate whether there is a pure strategy equilibrium, as McCubbins and Schwartz claim, or whether there is a mixed strategy equilibrium, as Aberbach's empirical findings seem to indicate. Even further, the model will help discriminate between the necessary and sufficient conditions for each case.

All the above examples have to do with questions of authority. What is the appropriate strategy to induce citizens, firms, countries, subordinates, bureaucracies, and so on to conform to a set of rules? The results are quite different from the conventional wisdom. Conformity to the rules cannot be achieved (at equilibrium) either by the carrot or by the stick. The impact of such measures will be the reduction of monitoring frequencies. If the frequency of deviant behavior is the target, the payoffs to be modified are those of the monitoring agency.

What explains these discrepancies between widely held beliefs and closer game-theoretic scrutiny? At the heart of the matter lies the concept of probability. The very concept of probability in the social sciences involves the mixture of two completely different concepts, which I will call partition frequencies and mixed strategies.

Partition frequencies refers to some unknown or unknowable desirable element of information that enables the decision maker (or the player) to make the correct choice in a particular situation. Therefore, this information is a public good. An actor who is willing to provide it will provide substantial benefits not only to him- or herself but to all other actors involved in the same or a similar situation. *Mixed strategies,* on the contrary, are probabilities generated by rational players in order to disseminate information about the choice of their strategies. In fact, the opponent who acquires information about the choice of strategies of a player will have a decisive advantage over this player. Therefore, information about mixed strategies is a private good that cannot be acquired from the opponent. The concept of a mixed strategy Nash equilibrium is designed to prevent the opponent from acquiring information that would put him or her in a strategically advantageous position. Thus, the second use of the word *probability* is exactly the opposite of the first.

The distinction between partition frequencies and mixed strategies is meaningful only in the social sciences. The Robinson Crusoe fallacy is the substitution of partition frequencies for mixed strategies, and this is the culprit causing the confusion and wrong conclusions in all the previous examples.

Consider the meaning of the word *probability* in various kinds of explanations of social phenomena. Elster (1983) claims that there are three possible types of explanation in the social sciences:

causal, functional, and intentional. He immediately proceeds one step further by questioning the validity of functional explanations in the social sciences.[12] Regardless of their validity, functional explanations have been shown to be shorthand expressions for causal arguments (Hempel 1965, chap. 11). Therefore, we are left with two different generic explanations: causal and intentional. The difference is that causal explanations explain phenomena by some set of antecedent conditions, while intentional explanations make use of consequences and conscious decision makers (or players).

I will focus on intentional explanations because they are the only kind of explanations in which the distinction between partition frequencies and mixed strategies is meaningful. Intentional explanations are divided into decision-theoretic and game-theoretic species. The difference between these two categories is that decision theory situations use what Elster calls "parametric rationality," meaning that the decision maker can anticipate (exactly in decision theory under certainty, probabilistically in decision theory under risk) the situation that will pertain. In game theory, however, all players respond to each others' moves optimally, adapting to a changing situation in which the changes are caused in part by the players' own moves.

In decision theory, if one of the parameters of the situation changes, this change may create a different choice situation and therefore a different outcome. Thus, the probabilities of different choices correspond to the probabilities of different choice situations. Probabilities in decision-theoretic situations are, therefore, partition frequencies.

I will subdivide game-theoretic problems into three different classes: (1) dominance solvable games and games with unique, pure strategy equilibria, (2) games with multiple, pure strategy equilibria, and (3) games without pure strategy equilibria (that have one, mixed strategy equilibrium) (Nash 1951).

In dominance solvable games, each one of the players can eliminate his or her dominated strategies; when all players successively follow this procedure several times, if necessary, there remains only one acceptable strategy for each player (Moulin 1981). Similarly, if assuming perfectly rational players (more precisely, if perfect rationality is common knowledge) in games with a unique pure strategy equilibrium, one expects players to choose the strategies that correspond to this unique equilibrium because they know that these are the only mutually optimal responses.[13] In such games the outcome is as predictable as the apple falling on Newton's head. If a mixture of outcomes is observed, the logical inference to be made is that there is a mixture of games.[14] Probabilities in such cases are partition frequencies.

The case of multiple equilibria is more complicated because a unique set of mutually optimal strategies does not exist. Here, the players face a coordination problem: each one may have different preferences about which is the most desirable equilibrium. However, each prefers to choose the appropriate equilibrium strategy, as all the others do. As a result, there is created an incentive for all of them to sort the different kinds of equilibria by using, for example, a common signal. If such a signal exists, it will be used and the situation will be simplified into a decision problem of the previous category. The presence of the signal will indicate the choice of a specific set of strategies and the frequency of the signal will indicate the probability of choice of the particular equilibrium. I do not know of any account of such games in which the choice of equilibria is made endogenously, without the use of some external signal.

In the case of no pure strategy equilibrium, each player has a unique mixed

Table 4. Different Problems and the Relevant Concept of Probability

Nature of Explanation	Source of Probabilities	Kind of Probabilities
Causal	mixture of causes	partition frequencies
Intentional		
Decisions	mixture of situations	partition frequencies
Games with more than one equilibrium	different strategy combinations	partition frequencies
Games with one pure strategy equilibrium	mixture of games	partition frequencies
Games with one mixed strategy equilibrium	mixed strategies	mixed strategies

equilibrium strategy (a probability distribution over his or her pure strategies) that is the optimal response to the mixed strategy choices of the others. Deviation from these strategies will provoke an infinite cycle of deviations in which each player either will be punished and made worse off or find an opportunity to deviate further. Therefore, each player has an incentive to prevent the opponent(s) from guessing how that player will choose. This can be achieved by randomizing the strategies. In the absence of a pure strategy equilibrium, therefore, each player will want to hide his or her choice. Probabilities in this case express an element of strategic choice, not an "objective" situation that would have been a means of exploitation of some players by others.

Table 4 summarizes the argument. Causal explanations have been included for reasons of completeness. Without entering into the important epistemological debate between determinists and indeterminists,[15] it can be said that whenever the concept of probabilistic outcomes occurs in causal explanations, it refers to the mixture of different causes and therefore expresses partition frequencies.

To summarize, there are two different ways the word *probability* can be used. I have called them partition frequencies and mixed strategies. Mixed strategies do not occur in causal arguments, in decision-theoretic situations, or even in games with pure strategy equilibria. They occur only in games with mixed strategy equilibria so that a player will not be outguessed and bested by the opponent. I have provided several examples of such cases and think that they can be multiplied: any situation in which inside information is valuable and "intelligence" is used to collect it is prima facie a candidate to be modeled as a game without a pure strategy equilibrium in which probabilities are in fact mixed strategies and not partition frequencies.

Whenever such a substitution of partition frequencies for mixed strategies occurs, we are led to wrong beliefs. We expect a change in the payoffs of one player to influence that player's behavior, when in reality such a modification gets completely absorbed by a change in the strategy of the opponent.

Notes

I would like to thank Robert Bates, James DeNardo, Barbara Geddes, Clark Gibson, Bernie Grofman, Virginia Hauffler, Peter Lange, Thomas Schwartz, Kenneth Shepsle, and Michael Wallerstein for many useful comments. Financial support of the Academic Senate of UCLA is gratefully acknowledged.

1. The "sure thing" principle formally states that if two acts are such that the first is preferred over the second, *given a state of the world,* but have the same consequences otherwise, the first is *unconditionally* preferred over the second. This axiom does not hold if the states of the world are dependent on the acts chosen.

2. The reader may object that these statements assume both the public and the police to be unified actors. In this case it is true that what the police does depends on the action of the public. However—the argument may go—the behavior of the police does not depend on the actions of any individual member of the public. For the time being assume the public as a unified actor. I will address this objection in more detail when I present alternative specifications of the model leading to the same outcome.

3. It is easy to verify that the first derivatives of equation 2 with respect to d_1 and c_1 are negative. A more intuitive way to verify theorem 5 is to observe that a decrease in the numerator and the denominator of a fraction by the same amount decreases the fraction.

4. For a model with incomplete information and the proof of this proposition, see Tsebelis 1987c. For the general proof that a mixed strategy equilibrium can be supported by incomplete information assumptions, see Harsanyi 1973.

5. See Tsebelis 1987a. I am grateful to Michael Wallerstein for indicating this to me.

6. See Tsebelis 1987a for references to the empirical literature.

7. This is not the only possible assumption. One can assume, alternatively, that they will violate the law anyway, taking their chances in the subsequent litigation process. In this case a pure strategy equilibrium exists (*violate* for the industry and presumably *enforce* for the agency).

8. See Tsebelis 1987b for an extended development of the regulation game, as well as for a review of the economic literature.

9. See the review by Olson (1979).

10. See the exhaustive empirical study by Hufbauer and Schott (1985).

11. See the exhaustive empirical study by Hufbauer and Schott (1985).

12. See also Elster 1985. For a different opinion, see Stinchcombe 1968.

13. The same argument applies if there are multiple (Nash) equilibria all but one of which can be eliminated as unstable. In this case some refinement of the Nash equilibrium concept like perfect (Selten 1975), proper (Myerson 1978), or sequential (Kreps and Wilson 1982), can help eliminate "unreasonable" equilibria. For a full account see Damme 1983.

14. Or, of course, it might be that the assumption of common knowledge of rationality is wrong. In this case, the game can be modeled as a game of incomplete information (Kreps and Wilson 1982) or imperfect information (Trockel 1986). These cases do not present a unique equilibrium and therefore belong to the next category.

15. For a deterministic view of causality, see Bunge 1979. For an indeterministic view, see Popper 1965.

References

Aberbach, Joel D. 1987. "The Congressional Committee Intelligence System: Information, Oversight, and Change." *Congress and the Presidency* 14:51–76.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76:169–217.

Bunge, Mario. 1979. *Causality and Modern Science.* 3d ed. New York: Dover.

Damme, Eric C. van. 1983. *Refinements of the Nash Equilibrium Concept.* New York: Springer-Verlag.

Eells, Ellery. 1982. *Rational Decision and Causality.* New York: Cambridge University Press.

Ehrlich, Isaac. 1973. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy* 81:521–64.

Ehrlich, Isaac, and George D. Brower. 1987. "On the Issue of Causality in the Economic Model of Crime and Law Enforcement: Some Theoretical Considerations and Experimental Evidence." *American Economic Review* 77:99–106.

Elster, Jon. 1983. *Explaining Technical Change.* New York: Cambridge University Press.

Elster, Jon. 1985. *Making Sense of Marx.* New York: Cambridge University Press.

Harsanyi, John C. 1973. "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points." *International Journal of Game Theory* 2:1–23.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York: Free Press.

Hufbauer, Gary C., and Jeffrey J. Schott. 1985. *Economic Sanctions Reconsidered.* Washington: Institute for International Economics.

Kohlberg, Elon, and Jean-François Mertens. 1986. "On the Strategic Stability of Equilibria." *Econometrica* 54:1003–37.

Kreps, David M., and Robert Wilson. 1982. "Sequential Equilibria." *Econometrica* 50:863–94.

Luce, Robert Duncan, and Howard Raiffa. 1957. *Games and Decisions.* New York: John Wiley & Sons.

McCubbins, Mathew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols and Fire Alarms." *American Journal of Political Science* 28:165–79.

Moulin, Herve. 1981. *Game Theory for the Social Sciences.* New York: New York University Press.

Myerson, Roger B. 1978. "Refinements of the Nash Equilibrium Concept." *International Journal of Game Theory* 7:73–80.

Nash, John F., Jr. 1951. "Noncooperative Games." *Annals of Mathematics* 54:289–95.

Nagin, Daniel. 1978. "General Deterrence: A Re-

view of the Empirical Evidence." *Deterrence and Incapacitation*, ed. Alfred Blumstein, Jacqueline Cohen, and Daniel Nagin. Washington: National Academy of Sciences.
Olson, Richard S. 1979. "Economic Coercion in World Politics." *World Politics* 31:471–94.
Ordeshook, Peter C. 1986. *Game Theory and Political Theory*. New York: Cambridge University Press.
Popper, Karl R. 1965. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper & Row.
Savage, Leonard. 1954. *The Foundations of Statistics*. New York: Dover.
Selten, Reinhard. 1975. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory* 4:25–55.
Stigler, George J. 1970. "The Optimum Enforcement of Laws." *Journal of Political Economy* 78: 526–36.
Stinchcombe, Arthur. 1968. *Constructing Social Theories*. New York: Harcourt, Brace & World.
Tittle, Charles R. 1983. "Social Class and Criminal Behavior: A Critique of the Theoretical Foundation." *Social Forces* 62:334–58.
Trockel, Walter. 1986. "The Chain-Store Paradox Revisited." *Theory and Decision* 21:163–79.
Tsebelis, George. 1987a. "Penalty Has No Impact on Crime: A Game-theoretic Analysis." International Political Economy Working Paper no. 23. Duke University.
Tsebelis, George. 1987b. "The Impact of Fines on Regulated Industries: Game Theory vs. Decision Theory." International Political Economy Working Paper no. 25. Duke University.
Tsebelis, George. 1987c. "Are Sanctions Effective? A Game-theoretic Analysis." International Political Economy Working Paper no. 24. Duke University.

George Tsebelis is Assistant Professor of Political Science, University of California, Los Angeles, CA 90024.