Inducible Games: Using Tit-for-Tat to Stabilize Outcomes

Brams, Steven J. and Kilgour, D. Marc

October 2012

Inducible Games:
Using Tit-for-Tat to Stabilize Outcomes

Steven J. Brams
Department of Politics
New York University
New York, NY 10012
USA
steven.brams@nyu.edu

D. Marc Kilgour
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5
CANADA
mkilgour@wlu.ca

October 2012

Abstract

Assume it is known that one player in a 2×2 game can detect the strategy choice of its opponent with some probability before play commences. We formulate conditions under which the detector can, by credibly committing to a strategy of probabilistic tit-for-tat (based on its imperfect detector), induce an outcome favorable to itself. A non-Nash, Pareto-optimal outcome is *inducible*—that is, it can be stabilized via probabilistic tit-for-tat—in 20 of the 57 distinct 2×2 strict ordinal games without a mutually best outcome (35 percent). Sometimes the inducement is "weak," but more often it is "strong." As a case study, we consider the current conflict between Israel and Iran over Iran's possible development of nuclear weapons and show that Israel's credible commitment to probabilistic tit-for-tat can, with sufficiently accurate intelligence, induce a cooperative choice by Iran in one but not the other of two plausible games that model this conflict.

Inducible Games: Using Tit-for-Tat to Stabilize Outcomes[1]

1. Introduction

A generation ago Axelrod (1984) argued that *tit-for-tat*—"I'll cooperate if you do; otherwise, I won't"—is a robust strategy for inducing cooperation in *repeated* Prisoners' Dilemma (PD) games with an uncertain end. Among his findings were that tit-for-tat defeated all other strategies proposed in two tournaments. Laboratory experiments of repeated PD conducted over several decades have shown that tit-for-tat often induces cooperation, even though players have strictly dominant strategies of not cooperating in any single play. In addition, tit-for-tat seems to explain how cooperation emerges in several real-life situations.

But there have been numerous criticisms of Axelrod's conclusions. Some, for example, have argued that tit-for-tat is brittle strategy, because if players switch to defection if an opponent defects, then they will be locked into defection until the end of play. By contrast, a delayed or more forgiving strategy (e.g., defect only if an opponent defects twice in a row, or probabilistic forms of tit-for-tat that render it more generous) avoids this problem and better allows for mistakes (Molander, 1985). Other critics have pointed out that still different strategies, including those that allow for mutations and different kinds of reciprocity in an evolutionary setting, frequently work better (Nowak, 2006, 2011; Sigmund, 2010). On one matter, however, there is a consensus: No strategy is optimal in all situations, which in our view leaves *all* prescriptions about how to act in repeated PD suspect.

[1] We thank Bryan Bruns for valuable comments on an earlier version of this paper.

In this paper, we do not propose a new strategy for repeated PD or any other specific games, nor do we analyze the dynamics of the evolution of cooperation using evolutionary game theory. Instead, we ask whether tit-for-tat can be put into a simultaneous-choice framework (repeated play and evolutionary game theory assume sequential play), wherein one player has imperfect, or probabilistic, information about its opponent's action *in advance*. In the single play of a game, this player commits to choosing its strategy based on this imperfect information. The opponent is aware of this commitment, regards it as credible, and responds rationally to it.

We apply this framework to all 2×2 games of conflict, of which PD is just one.[2] We identify in which games this commitment, if believed, can induce an outcome favorable to the detecting player if its detector is sufficiently accurate.

We note that it is common for players to make pledges about the actions they intend to take. For instance, the doctrine of mutually assured destruction (MAD) was essentially a reciprocal pledge by the two superpowers to abide by tit-for-tat, which was effective in preventing nuclear war over the nearly 50-year history of the Cold War.

In the current model, only one player makes a commitment, though it does not matter which one if the game is symmetric, like PD. As we will show, this commitment undercuts the dominance of noncooperation in one-shot PD, rendering cooperation on the part of *both* players—in order to maximize their expected payoffs— optimal, given that the detector's probability of correct detection is sufficiently high.

---

[2] PD is one of exactly 78 distinct 2×2 strict ordinal games in which two players, each with two strategies, can strictly rank the resulting four outcomes from best to worst. When the 21 games with a mutually best outcome are excluded, there remain 57 games, called *conflict games*, which include PD. There has been considerable recent work on classifying 2×2 strict ordinal games and understanding their topology. See Robinson and Goforth (2005), Bruns (2011), and citations therein.

To be sure, giving the detector prior knowledge and the ability to make a credible commitment alters the structure of the classic PD. But the point of our analysis is to show that this reformulation, which we consider realistic, can lead the players to a preferred outcome (mutual cooperation), in PD and other games, that otherwise would elude them because it is unstable.

During the Cold War, each superpower assiduously sought to acquire information about the likely strategy choice of its opponent. A state's intelligence is typically assembled from a variety of sources, including the interception of electronic signals (possibly through computer hacking), observations from manned or unmanned overflights, satellite reconnaissance, and human informers and spies. Unsurprisingly, improvements in the ability to ferret out information about an opponent's plans have led to corresponding improvements in the ability to conceal or manipulate this information, so what is gleaned from sophisticated intelligence operations today may be no more accurate or reliable than when espionage relied solely on human intelligence.

There is no denying the enormous resources now devoted to intelligence. Since the early 1960s, the most significant improvement in the detection capabilities of states has come from the use of reconnaissance satellites, which were first deployed by the superpowers but are now widely used. President Lyndon Johnson claimed that space reconnaissance had saved enough in military expenditures to pay for all other military and space programs (Biddle, 1972). President Jimmy Carter, in the first public acknowledgment of photo reconnaissance satellites, said that "in the monitoring of arms control agreements, they make an immense contribution to the security of all nations" (*Chicago Tribune*, October 2, 1978, p. 2).

Although intelligence is not always accurate, a maxim of intelligence agencies is that some information, even if imperfect, is better than none. We do not address the problem of acquiring high-quality information but ask, instead, when and how less-than-perfect information can be used to advantage by players trying to influence the behavior of opponents.

In this paper, we suppose that one player (the *detector*) can detect with some probability the strategy choice of its opponent. *Probabilistic tit-for-tat* is a pledge by the detector (the *inducer*) to cooperate if and only if it detects that its opponent (the *inducee*) also cooperates.

We assume throughout that the resulting cooperative outcome is *not* a Nash equilibrium.[3] If it were, it could be induced by the inducer's commitment to choose a particular strategy (with certainty), rather than a contingent commitment based on what it detects. Our goal is to identify those games in which probabilistic tit-for-tat is critical in stabilizing a non-Nash cooperative outcome.

Our main finding is that probabilistic tit-for-tat can induce a *favorable outcome*—better for the detector than any pure-strategy Nash equilibrium—in 20 of the 57 conflict games (35 percent). We call games in which tit-for-tat can induce such an outcome *inducible games*.

---

[3] Technically, a *Nash equilibrium* is a pair of strategies, associated with an outcome, from which neither player would unilaterally depart because it would not benefit by doing so. But because each outcome in a strict ordinal game is associated with a unique pair of strategies, these outcomes can be used to define strategy pairs that are in equilibrium. We consider in this paper only pure-strategy Nash equilibria, but we note that sometimes it is possible for mixed-strategy equilibria to be perturbed slightly to induce certain outcomes (Brams and Kilgour, 1988).

While the inducer, by committing to probabilistic tit-for-tat, renders it rational for the inducee to choose its cooperative strategy,[4] the inducer's information may be inaccurate. Hence, it may not always be rational for the inducer to choose cooperation when it detects that the inducee will cooperate.

But if the inducer's detector is sufficiently accurate, it benefits from inducing cooperation, via probabilistic tit-for-tat, in inducible games. For the inducee, these include games in which it obtains a rather poor (i.e., next-worst) outcome, though in most inducible games the inducee does better (i.e., obtains a next-best outcome) than it would at a Nash equilibrium.

The paper proceeds as follows. In section 2, we describe our model of inducement. In section 3 we specify the calculus of the inducee, which we assume to be Row, and in section 4 the calculus of the inducer (Column). These specifications, which are conditions on the game being played, allow for three payoff configurations for Row, and eight for Column, producing 24 potentially inducible games (two of them are the same game, with the inducer and inducee interchanging roles). We distinguish weakly from strongly inducible games, wherein inducement is more reliant on the quality of the detector.

In section 5 we discard one of the duplicative games, and three others, because they have Nash equilibria that Column (the inducer) prefers to the inducible outcome. In these games, the inducer can effect a Nash equilibrium by a commitment strategy simpler that probabilistic tit-for-tat. This leaves 20 inducible games, in which the cooperative outcome that Column can induce benefits it and, usually to a lesser extent, Row.

---

[4] Although the strategy "cooperate" is well defined in PD, it is not always clear what it means in other games, which is why we associate it with a favorable outcome for the inducer. As we will see, it may be either a next-best or next-worst outcome for the inducee.

In section 6 we apply our analysis to international relations, focusing on the current conflict between Iran and Israel over Iran's possible development of nuclear weapons and Israel's choice of attacking or not attacking Iran's nuclear facilities. We summarize our analysis, and draw several conclusions, in section 7.

2. Imperfect Detection and Probabilistic Tit-for-Tat

Assume Row and Column play a 2×2 game, depicted in Figure 1, wherein each player can choose either to cooperate (C) or defect (D). Assume further that Column has a detector, which signals the choice that Row is about to make (C_R or D_R).

The probability that the detector gives a correct (accurate) signal depends on whether Row intends to choose C_R or D_R. We define the quality of the detector as two conditional probabilities,

p = Pr {detector signals C_R | Row chooses C_R}

q = Pr {detector signals D_R | Row chooses D_R},

where p and q are parameters satisfying $0 \leq p, q \leq 1$.

Figure 1 about here

We assume the values of p and q are *common knowledge*: Each player knows these values, knows that the other player knows them, knows that this knowledge is known, ad infinitum.[5] However, only Column acts on the basis of the signal it receives; Row acts on

[5] Such common knowledge might be acquired by Column's demonstration of the quality of its detection equipment to Row. In a related model of arms-control inspection, putting an inspectee "on notice" via such a prior move can induce greater cooperation on the part of the inspectee (Brams and Kilgour, 1992). Of course, in other situations an inspector may wish to hide its ability to detect the choice of an opponent, lest the opponent take countermeasures to conceal its choices. In section 5 we show that there are a few games

the presumption that Column chooses its action according to its commitment. For example, if Row chooses C_R, then Column's choice will depend on whether the detector signals C_R (with probability p) or D_R (with probability $1-p$).

Our model of the detection process is quite general. It might involve specific, fixed values of p and q. Alternatively, Column might be faced with a so-called *characteristic curve* $q = f(p)$ such that, for each value of p, $0 \le p \le 1$, the greatest value of q that can be achieved is $f(p)$. Column chooses a value of p, say p_0, and then uses the detector with characteristic $(p_0, f(p_0))$. It is usually assumed that $f(0) = 1, f(1) = 0, f(p) > p$ for $0 < p < 1$, and $f(p)$ is strictly decreasing in p.[6] A typical characteristic curve is shown in Figure 2. Note that the point (p^*, p^*) is the point where the curve crosses the 45-degree line defined by $p = q$; at all other points of the curve, q is either less than or greater than p. The distance of (p^*, p^*) from (1, 1) is a measure of the degree of imperfection of the detector.

Figure 2 about here

How should Column use its detector in the play of a 2×2 game? We assume that it conditions its choice, *C* or *D*,[7] on the signal that it receives from its detector. More specifically, if Column wishes to induce *CC*[8]—subject to the limitations of its detector—it will respond to Row's (apparent) choice of *C* with *C*, and Row's (apparent) choice of *D* with *D*.

---

in which, after Row chooses to cooperate, either Column or Row can benefit when Column misdetects Row's choice and chooses not to cooperate.

[6] See Lindgren (1976). For an application to arms races, see Brams, Davis, and Straffin (1979a, 1979b) and Dacey (1979).

[7] We drop the subscript of the player (*C* or *R*) that chooses a strategy when its identity is clear from the context.

[8] When a strategy pair *XY* is used to identify an outcome, *X* is Row's strategy and *Y* is Column's strategy.

When Row chooses *C* and Column incorrectly detects *D*, which occurs with probability 1–*p*, a type 1 error (a false negative) occurs; when Row chooses *D* and Column incorrectly detects *C*, which occurs with probability 1–*q*, the error is of type 2 (a false positive). How damaging each type of error is will depend on the situation, which we will say more about later.

Henceforth, we assume that Column credibly commits to probabilistic tit-for-tat, pledging that "I'll cooperate if I detect that you will cooperate; otherwise, I won't." Our objective is to identify the 2×2 games in which there is a cooperative outcome (*CC*), which is not stable on it own (i.e., is not Nash). But it can be stabilized by probabilistic tit-for-tat, providing the detector is good enough (even if less than perfect). Among other things, we show that *CC* must be at least as favorable to Column (the inducer) as to Row (the inducee). Whereas *CC* is always Column's best or its next-best outcome, it is never best for Row and may even be Row's next-worst outcome.

3. Row's (Inducee's) Calculus

We assume that Column can correctly detect Row's choice of *C* with probability *p*; similarly, it can correctly detect Row's choice of *D* with probability *q*. Given that Column's commitment to probabilistic tit-for-tat is credible, Row's expected payoffs from its choices of *C* and *D* are as follows:

$$E_R(C) = pa_{11} + (1–p)a_{12}$$

$$E_R(D) = qa_{22} + (1–q)a_{21}.$$

Row prefers to choose *C* iff $E_R(C) \geq E_R(D)$, which is true iff

$$p(a_{11} – a_{12}) + q(a_{21} – a_{22}) \geq a_{21} – a_{12}. \qquad (1)$$

We call (1) the *Inducement Condition for Row* and next specify several reasonable conditions which, if satisfied, justify the conclusion that the existence of the detector, and Column's credible commitment to rely on it, can induce Row to choose *C*.

First, it must be the case that

$$a_{21} > a_{11}, \tag{R1}$$

because otherwise a detector is not necessary—Column can induce Row to choose *C* simply by credibly committing to choose *C*. This follows from the fact that if (R1) fails, Row—knowing that the outcome will be in the first column if Column commits to *C*—will choose *C* because $a_{11} \geq a_{21}$, obviating the need for detection.

To understand our second condition, suppose that Row believes that Column's detector works perfectly, so Column will always choose *C* in response to Row's choice of *C*, and *D* in response to Row's choice of *D*. Then, effectively, Row must choose between *CC* and *DD*. To induce *CC*, Row must find it no less preferable than *DD*, that is

$$a_{11} \geq a_{22}, \tag{R2}$$

because otherwise Row would prefer that Column actually carry out its threat—choose *D* after detecting *D*.

Notice that (R2) is equivalent to the Inducement Condition for Row, (1), in the case that $p = q = 1$. Thus, if (R2) holds, inducement always succeeds if the detector is perfect.

Our third condition, which we call *inducibility*, is roughly that inducement depends on the detector's being good enough. This condition allows for inducement when the detector is less than perfect.

We distinguish two kinds of inducibility:

1. Inducement is *weak* when it *fails* for the poorest possible detector, which is one with $(p, q) = (0, 0)$.

2. Inducement is *strong* if, when it works for some detector, then it also works for any detector that is an improvement on the first.

Improving a detector means increasing its values of p, q, or both. Thus, for strong inducibility, we require that if a detector works for a detector with characteristic (p, q), then it also works for a detector with characteristic (p', q'), provided that $p' \geq p$ and $q' \geq q$.

We show in the Appendix that a necessary condition for weak inducibility is

$$a_{12} < a_{21}, \tag{R3}$$

and a necessary condition for strong inducibility is

$$a_{12} \leq a_{11}. \tag{R4}$$

For 2×2 strict ordinal games, assumptions (R1) and (R2) imply that $a_{22} < a_{11} < a_{21}$. Because Row's payoff, a_{12}, can be inserted at any point in this ordering, there are four possible strict orderings for Row:

(a) $a_{12} < a_{22} < a_{11} < a_{21}$.

(b) $a_{22} < a_{12} < a_{11} < a_{21}$.

(c) $a_{22} < a_{11} < a_{12} < a_{21}$.

(d) $a_{22} < a_{11} < a_{21} < a_{12}$.

These four orderings can be conveniently described by setting the utilities of Row's best, next best, next worst, and worst outcomes equal to 4, 3, 2, and 1, respectively, which correspond to (a) $a_{12} = 1$, (b) $a_{12} = 2$, (c) $a_{12} = 3$, and (d) $a_{12} = 4$. Note that (R3) holds in cases (a), (b), and (c), whereas (R4) holds only in cases (a) and (b). Thus, cases (a) and (b) are strongly (and also weakly) inducible, whereas case (c) is weakly but not strongly inducible, and case (d) is not inducible.

The shaded regions in the (p, q) unit squares of Figure 3 show the characteristics of all detectors for which the Inducement Condition for Row, (1), holds. These regions, separated by straight lines from the noninducibility (unshaded) regions, are calculated as if the aforementioned ordinal values (4, 3, 2, 1) were cardinal.

Figure 3 about here

In the Appendix, we draw a distinction between the strongly inducible games described by Figures 3(a) and 3(b). Assume a detector has a characteristic curve of the kind shown in Figure 2. Then Column can choose a detector characteristic, $(p_0, f(p_0))$, that satisfies the Inducement Condition for Row, (1), in Figure 3(b), whereas this condition may fail in Figure 3(a) for all possible detectors. We say that strong inducement is *guaranteed* in Figure 3(b) but not in Figure 3(a); for details, see the Appendix.

4. Column's (Inducer's) Calculus and 24 Potentially Inducible Games

Having determined the conditions under which Row can be induced to choose *C* when Column uses an imperfect detector and makes a credible commitment to

probabilistic tit-for-tat, we next inquire under what conditions this commitment would benefit Column as inducer.

If Column's detector were perfect, recall from section 3 that it offers, in effect, Row the choice between *CC* and *DD*. If Column wishes to induce *CC*, then it must weakly prefer it to *DD*:

$$b_{11} \geq b_{22}. \tag{C1}$$

In the presence of (R2), this means that *CC* is *Pareto-superior* to *DD*—both Column and Row agree that *CC* is preferable to *DD*.

Now assume that conditions (R1) – (R3) hold so that, as noted earlier, a_{21} is Row's best payoff. If it were the case that $b_{21} \geq b_{11}$, then *DC* would be a Nash equilibrium, and would be preferred by both players to *CC*. In such a case, it would make no sense for Column to try to induce *DC*. We therefore require that

$$b_{11} > b_{21}, \tag{C2}$$

a condition that implies that the two players have opposite preferences for *CC* and *DC*—Column prefers *CC*, Row *DC*—which is precisely why a sufficiently good detector is needed to induce *CC* via tit-for-tat.

In section 3, we noted that there are three configurations of Row's payoffs – denoted (a), (b), and (c) – that are consistent with (R1) – (R3). It is easy to verify that eight configurations of Column's payoffs are consistent with conditions (C1) and (C2) (see Figure 4). It follows that $8 \times 3 = 24$ games, shown in Figure 4, satisfy all these conditions. As noted in Figure 4, one-third of these games are weakly but not strongly

inducible, whereas the rest are both weakly and strongly inducible. But, we caution, the 24 games are only "potentially" inducible, for reasons we will discuss in section 5.

Figure 4 about here

Recall from section 1 (ftn. 2) that there are 57 2×2 conflict games (for a complete listing, see the Appendices in Brams, 1994, 2011). Each of the listed games is equivalent to up to seven others, obtained by interchanging the players and/or their column or row strategies.

In fact, 23 of the 24 games in Figure 4 are not equivalent in this sense, making them truly distinct. However, #24i ('i' is for interchange) is not distinct: It is the mirror image of #24 (also potentially inducible), which is obtained by interchanging Row and Column in #24.

Thus, either player can induce *CC* in #24. Also worth noting is that two of the potentially inducible games in Figure 4, PD (#32) and Chicken (#57), are symmetric, so the strategic problems faced by their players are identical. Hence, in these games as well, either player can induce *CC*.

5. 20 Inducible Games

In four of the 24 games in Figure 4, Column will not in fact be motivated to induce *CC*, because there is a pure-strategy Nash equilibrium that Column prefers to *CC*. Because Column can effect these equilibria by credibly committing to choosing its strategy associated with them, it gains no benefit in committing to tit-for-tat and relying on an imperfect detector.

We, therefore, remove these four games (#20, #22, #53, and #57) from the list of 24. In two of them (#20 and #53), the outcome that Column can induce with a detector, (2,3), is in fact Pareto-inferior to the Nash equilibrium (3,4)— worse for both Column and Row. In game #53, there is a second Nash equilibrium, (4,2), which Row prefers to (3,4) and could, therefore, effect by credibly committing to its strategy of *D*.

In Chicken (#57), either player can stabilize *CC*, using probabilistic tit-for-tat. But this game has two equilibria, each preferred by one player, so again there is the potential for a clash of commitments.

Eliminating the four aforementioned games from the potentially inducible games in Figure 4, we focus on the 2×2 conflict games in which there is an unstable *CC* outcome that can be stabilized using probabilistic tit-for-tat. In Figure 5, we classify these 20 inducible games into five classes, depending on their properties.

Figure 5 about here

Perhaps surprisingly, inducible games have a substantial overlap with the class of *difficult games* identified by Brams and Kilgour (2009); see also Brams (2011, ch. 5). In these games, *CC* is the unique outcome that is either best (4) or next best (3) for each player, is Pareto-superior to *DD*, but is not a Nash equilibrium. Brams and Kilgour (2009) showed that in a difficult game *CC* can be stabilized by a process tied to voting.

The difficult games can be viewed as a class that generalizes both Prisoners' Dilemma and Chicken. Remarkably, all 11 difficult games are potentially inducible; only Chicken is not inducible.

The classification shown in Figure 5 extends the classification of difficult games suggested by Brams and Kilgour (2009), which reflects their Nash equilibrium properties.

The 10 difficult games that are inducible fall into classes 1, 2, and 3 below. The complete classification is as follows:

1. The game is difficult and the *DD* outcome is a Nash equilibrium (4 games).
2. The game is difficult and there is a Nash equilibrium, but it is neither *CC* nor *DD* (3 games).
3. The game is difficult and there is no Nash equilibrium in pure strategies (3 games).
4. The game is not difficult, and the *CC* outcome is next-best for Row (4 games).
5. The game is not difficult, and the *CC* outcome is next-worst for Row (6 games).

It might seem a misnomer to call the (2,4) outcomes in the class 5 games "cooperative." In fact, these are exactly the six weakly inducible games. In contrast to the 14 strongly inducible games, wherein Row (the inducee) receives its next-best payoff (3) at *CC*, Row does relatively poorly in the class 5 games by receiving its next-worst outcome. Indeed, when Column misdetects and chooses *D* in these games, Row actually does better, obtaining its next-best payoff (3), a point we will return to later.

Of the 20 inducible games in Figure 5, only in games #3 (PD) and #5 is *CC* the next-best outcome for both players. In the 18 other inducible games, Column obtains its best outcome (4) and Row does not, suggesting that the inducer tends to do better.[9]

Our inducibility condition shows that inducement fails for a poor detector but succeeds for a good enough detector. A rough measure of how good a detector must be

[9] This statement is based on the comparative ranks of players, not on an interpersonal comparison of their utilities. It seems reasonable to claim that a player does better than its opponent when it obtains its best outcome and its opponent does not.

can be obtained from the *threshold probability for Row*, p_0, which is the least value of p such that (p, p) is within the inducement region. This calculation assumes that $p = q$.[10]

In Figure 3, the value $p = p_0$ defines the point where the 45-degree line enters the inducement region. By setting $q = p$ in (1), it is easy to verify that

$$p_0 = \frac{(a_{21} - a_{12})}{(a_{21} - a_{12}) + (a_{11} - a_{22})}. \qquad (2)$$

For a detector with characteristic on the 45-degree line, Row will be willing to choose *C* rather than *D* if and only if $p \geq p_0$. Because we are assuming (R2), it can be shown from (2) that $0 < p_0 < 1$ if and only if (R3) holds – that is, if and only if the game is weakly inducible.

To see the significance of the threshold value, p_0, suppose that the characteristic of a detector is known. If the value p^* defined earlier (see Figure 2) satisfies $p^* > p_0$, then the detector is good enough for inducement.[11] Of course, this threshold probability p_0—above which Column can induce Row to choose *C*, rendering *CC* possible if not likely—is calculated by assuming that the players' ordinal rankings for outcomes are cardinal utilities.

For specific inducible games, actual values of p_0 may be quite different from those shown in Figure 5. To illustrate, assume in game 9 that Row's cardinalization is 10, 9, 8, 1, so its worst payoff is far below its three others. Then applying (3), $p_0 = 1/5$, so even

[10] In the context of international relations, we discuss in section 6 when this assumption might be applicable and when not.

[11] The converse of this statement is false. Nonetheless, the value of p_0 is a useful measure of how good a detector must be for inducement to occur.

poor predictions by Column can still make it rational for Row to choose *C*, thereby precluding its worst outcome.

Paradoxically, in two of the inducible games, #3 (PD) and #5, in both of which *CC* yields (3,3), it is in Column's interest that p be greater than p_0—but not by much. In each game, Column benefits by misdetecting Row's choice of *C* on occasion, giving Column an "excuse" to choose *D* and thereby obtain its best outcome of (1,4). In these two inducible games, and only in these two, better predictions—above the threshold value of p_0—are *not* in Column's interest, whereas Row suffers its worst outcome when Column's detector fails and incorrectly signals Row's choice of *D*. In the more general formulation with detector characteristic (p, q), as defined in section 2, increases in p, beyond the minimum, are not in Column's interest.

Curiously, it is Row that benefits in the six class 5 games when Column's detector fails. In these games, Row obtains a payoff of 3 at *CD* instead of 2 at *CC*, whereas Column always does best (4) at *CC*. In these games, too, the players have opposite interests in the quality of detection, given that Column's detection probability is sufficient to induce Row to choose *C*.

These "opposite interests" of players in certain games, while anomalous, should not detract from our main result—namely, that in more than 1/3 of the 2×2 conflict games, Column can induce Row to make a choice favorable to Column. To the degree that Column's detector is accurate, this leads to an outcome that would otherwise be unstable. In section 6, we explore the implications of these results in international relations, using them to illuminate the strategic situation of Iran and Israel regarding Iran's possible development of nuclear weapons.

6. Tit-for-Tat in International Relations[12]

In the international arena, tit-for-tat depends on a state's (i) having intelligence on the probabilities of an opponent's choices, (ii) communicating to the opponent its intention to act according to this intelligence, and (iii) being believed by the opponent. The use of tit-for-tat presumes that a player not only can learn what its opponent is about to do but also can influence that opponent's choice, benefiting itself in the process. Our calculations demonstrate that in 20 of the 2×2 conflict games, the opponent can be induced to cooperate if the inducers's detection probability is sufficiently high, and this probability is known by both players.

For an opponent to respond cooperatively, however, requires that it believe the detector will follow tit-for-tat and do what its detector tells it to do, even though this might not be the detector's optimal strategy *after* inducement occurs. For this reason, the detector's reputation for keeping its word is crucial for probabilistic tit-for-tat to work.

As a case in point, consider the current conflict between Iran and Israel over the suspicions of Israel, as well as other countries and the International Atomic Energy Agency, that Iran is enriching uranium in order to develop nuclear weapons that could be used against Israel. Iran denies this intent, despite the discovery of previously hidden nuclear facilities and the uncovering of other deceptions; it claims, rather, that it desires to enrich uranium only as an alternative energy source to be used for civilian purposes.

Prime Minister Benjamin Netanyahu and other Israeli leaders have threatened to attack Iran and destroy its nuclear capability unless there is proof, based on a rigorous

---

[12] We thank Etel Solingen for valuable comments on an earlier version of this section. Her chapters on Iran and Israel in Solingen (2007) offer background on the past calculations of these countries' leaders about the acquisition of nuclear weapons. For more recent information on the effects of threats and sanctions on Iran, see Nader (2012).

inspection of its suspected nuclear facilities, that Iran is not developing nuclear weapons.. (Other Israeli leaders have opposed such an attack, arguing that at best it might delay but not stop Iran's acquisition of nuclear weapons.) At the time of writing (September 2012), Israel and Iran are at an impasse, with Iran denying international inspectors access to the facilities in question.

Because of its refusal, Iran has already suffered severe economic sanctions imposed by the United States, the European Union, and other countries. More sanctions are scheduled to go into effect. They all have a tit-for-tat aspect, with the sanctioners offering to relax or lift the sanctions if Iran agrees to allow inspections and credibly commits to stopping its efforts that could lead to the production of nuclear weapons. But other countries, including China and Russia, have opposed the use of sanctions.

The most immediate danger of armed conflict arises from Israel's threat to attack Iran's nuclear-production facilities. More specifically, Israel's position is that, failing an agreement, it will attack Iran's facilities before a point of no return—called a "zone of immunity" by Israeli Defense Minister Ehud Barak (Landler and Sanger, 2012)—is reached, when these facilities become sufficiently hardened (they are inside a mountain) to be effectively impregnable. Whether the United States would actively participate in such an attack, or covertly facilitate it, is unclear, but President Barack Obama said on March 8, 2012, that the United States "will always have Israel's back."

Israel has never publicly acknowledged possessing nuclear weapons, but is widely presumed to have them; it has said that it would not be the first party to introduce them into a conflict. Because the present Israeli government avers that Iran's acquisition of nuclear weapons threatens its existence, it seems ready to arrest Iran's development of

them if economic sanctions or covert actions—including assassinations and cyberwarfare, which have been carried out already (Bergman, 2012; Bronner, 2112; Erdbrink, 2012)—do not work.

Unlike the superpowers during the Cold War, Israel appears unwilling to rely on its own nuclear deterrent and MAD, perhaps in part because it fears that terrorists could gain control of any nuclear weapons Iran develops. Israel's small size makes its survival an issue—even if retaliation is possible—whereas Iran's ability to absorb a retaliatory strike is greater, possibly giving it an incentive to preempt with nuclear weapons.

In the two games shown in Figures 6a and 6b, Iran chooses between developing (*D*) or not developing ($\overline{D}$) nuclear weapons, and Israel chooses between attacking (*A*) or not attacking ($\overline{A}$) Iran's nuclear facilities.[13] Israel's preferences are the same in the two games, whereas Iran's preferences vary, with Iran's next-best and next-worst outcomes interchanged in the two games.

Figures 6a and 6b about here

We assume Israel's ranking to be $\overline{D}\ \overline{A} > DA > \overline{D}A > D\overline{A}$. As justification, there is little doubt that Israel would most prefer a cooperative solution ($\overline{D}\ \overline{A}$), in which Iran does not develop nuclear weapons so no attack is required, and least prefer that Iran develop nuclear weapons without making an effort to stop their production ($D\overline{A}$). Between attacking weapons that are being developed (*DA*) and mistakenly attacking weapons that are not being developed ($\overline{D}A$), we assume that Israel would prefer the

[13] Biran and Tauman (2008) analyze a game modeling a similar situation in which there is imperfect detection, but they do not distinguish between type 1 and type 2 errors. They prove several propositions relating the detection probability, which may or may not be common knowledge, to different equilibria in the game.

former (the latter strategy would create a crisis, but it would not be disastrous to Israel's security).

As for Iran, at least given its present leadership, we assume its most preferred outcome is to develop nuclear weapons without being attacked ($D\,\overline{A}$), and its least preferred is not to develop nuclear weapons and be attacked anyway ($\overline{D}\,A$). In between, Iran's preferences are less clear. In Figure 6a, we assume that Iran prefers the cooperative outcome ($\overline{D}\ \overline{A}$) to the noncooperative outcome (DA), and in Figure 6b we assume the reverse. Thus, the issue is whether Iran prefers to develop weapons and be attacked, or neither.

In both of our games, D is a dominant strategy for Iran, and the unique Nash equilibrium is the noncooperative outcome (DA). Figure 6a is the weakly and strongly inducible game #1 (Figure 5), while Figure 6b is not inducible. In this game, Iran (Row) does best (4) or next best (3) by choosing its dominant strategy, D, so it cannot be induced to cooperate, independent of the quality of Israel's detection apparatus.

By contrast, probabilistic tit-for-tat does induce Iran's choice of $\overline{D}$ in game #1 if Israel's detection capability is good enough (assuming the 4, 3, 2, 1 cardinalization of this game in Figure 5, the threshold condition is $p > p_0 =$ ¾ if $p = q$). But if Iran only barely prefers $\overline{D}\ \overline{A}$ to DA, so that the gap between a_{11} and a_{22} is much smaller than the gap between a_{21} and a_{12}, near-perfect detection would be required. For example, utilities of 4, 3, 3.1, 1 produce a threshold of $p_0 = 0.97$.[14]

The assumption that $p = q$ (see section 2) may be appropriate in this particular case. If Israel correctly detects that Iran is developing nuclear weapons, it will probably

[14] A related question is whether detection of uranium enrichment or actual weaponization, or something in between, would constitute a *casus belli* for Israel.

correctly detect when it is not, though proving a "negative" can be difficult.[15] To be sure, in other situations (e.g., medical testing), the probabilities of type 1 and type 2 errors may be designed to be different, but this seems unlikely in international relations when intelligence is unbiased.

As we noted earlier, Iran has not proved inducible so far, at least under the pressure of economic sanctions, though this could change. Whether Iran will be persuaded to be more forthcoming as sanctions are increased, or under the threat of imminent attack by Israel, may well depend on whether its preferences are closer to those in the Figure 6a or Figure 6b game. But we emphasize that only in the former game can Iran be induced to cooperate.

Even if Israel's detection probability is greater than the threshold value required for tit-for-tat to induce cooperation, there may be complicating factors. For example, if Israel's detection probability is not common knowledge, as we assumed earlier, Iran may believe that Israel's intelligence capabilities are insufficient to ascertain its development of nuclear weapons. So it may proceed to develop them, mistakenly thinking that its efforts will not be detected.

In summary, we have suggested that the current conflict between Iran and Israel over Iran's possible development of nuclear weapons can be represented by two plausible games. In one, Iran cannot be induced to cooperate via probabilistic tit-for-tat, no matter how accurate is the intelligence Israel has on its activities, but in the other it can be induced if Israel has a sufficiently high probability of correctly detecting Iran's strategy,

---

[15] So can proving a "positive," as the United States learned when it incorrectly detected the presence of weapons of mass destruction in Iraq before it launched its attack in March 2003. But intelligence in this case seems to have been biased by the Bush administration's extreme antagonism toward Saddam Hussein and its desire to depose him.

and Iran knows this. Clearly, the future remains uncertain, but we think our model clarifies the basis of this uncertainty.

7. Summary and Conclusions

We have shown that unstable (non-Nash) outcomes in 20 of the 57 2×2 games of conflict (35 percent) can be stabilized by probabilistic tit-for-tat. We called these outcomes "cooperative," because they are better for the detector (inducer) than any pure-strategy Nash equilibrium in these games.

While they are always best (4) or next-best (3) outcomes for the detector (inducer), they are either next-best (14 games) or next worst (six games) for the inducee. But even in the latter games (the class 5 games in Figure 5), no other outcome is Pareto-superior to the induced outcome.

In two of these games (#15 and #16), there are Nash equilibria that the inducee would prefer. Moreover, we showed that in a few games, if inducement works and the inducee chooses *C*, either the inducer or the inducee can benefit when the inducer misdetects the choice of *C* and responds with *D*.

In the ten difficult inducible games (classes 1, 2, and 3 in Figure 5), a group that includes Prisoners' Dilemma, the cooperative outcome is either best or next best for both players, and it is the only such outcome. In these games, the case is strongest that it is in the interest of both players to stabilize this outcome via probabilistic tit-for-tat.

In any such game, the quality (characteristic) of the detector must fall into the zone of inducement (Figure 3). If this characteristic is a point on a characteristic curve (Figure 2), the possibility of inducement depends on whether the characteristic curve enters the zone of inducement (details are given in the Appendix).

Our case study of the Iran-Israel conflict presented two games that plausibly model this conflict. One is both weakly and strongly inducible, whereas the other is not inducible. Time will tell whether this conflict is resolved peaceably; if it is, it seems likely that inducible game #1 was the one played, and Israel's use of probabilistic tit-for-tat was effective.

In conclusion, we have provided a version of tit-for-tat that is quite different from that used in repeated play or evolutionary game theory. We think this perspective offers new insight into how cooperation can be stabilized in a large number of games wherein it is known that one side has the capability, with sufficient accuracy, to detect and respond in a tit-for-tat manner to the strategy choice of its opponent.

Appendix

This Appendix shows how the definitions of weak and strong inducibility are connected to the conditions (R3) and (R4) utilitized in section 3. Weak inducement is distinguished from strong inducement as follows:

1. Inducement is *weak* when it *fails* for the poorest possible detector, which is one with $(p, q) = (0, 0)$.

2. Inducement is *strong* if, when it works for some detector, then improving the detector cannot cause it to fail: If inducement works for some (p, q), then it is strong when it works for every (p', q') such that $p' \geq p$ and $q' \geq q$.

To explore these conditions further, recall that

$$p(a_{11} - a_{12}) + q(a_{21} - a_{22}) \geq a_{21} - a_{12}, \tag{1}$$

which we called the *Inducement Condition for Row*, is a necessary condition for inducement. Assume that (R1) and (R2) hold, and note that, together, they imply that $a_{21} > a_{22}$, and therefore that the coefficient of q in the inducement condition is positive. It is then easy to verify that (1) holds for (p, q) with $p = 1$ and $q \geq q_1$ where $q_1 = \dfrac{a_{21} - a_{11}}{a_{21} - a_{22}}$. Note also that $0 < q_1 \leq 1$.

Similarly, it can be shown that (1) holds for (p, q) with $p = 0$ and $q \geq q_0$ where $q_0 = \dfrac{a_{21} - a_{12}}{a_{21} - a_{22}}$. The boundary that separates the values of (p, q) satisfying (1) from those values where (1) fails is the straight line joining $(0, q_0)$ and $(1, q_1)$. This line is shown in each panel of Figure 3. Observe that $q_0 > 1$ in Figure 3(a) and $q_0 < 0$ in Figure 3(d).

Clearly, it is always the case that, if inducement is possible for (p, q), then it is also possible for every (p, q') such that $q' \geq q$.

It is apparent from Figure 3 that weak inducibility fails in case 3(d), which arises whenever $q_0 \leq 0$, or

$$a_{12} \geq a_{21}.$$

Thus, a necessary condition for weak inducibility is

$$a_{21} > a_{12}. \tag{R3}$$

Put another way, (R3) means that *CC* can be stabilized using probabilistic tit-for-tat with the poorest possible detector. It is easy to verify that (R3) holds iff $q_0 \geq 0$, which is true in cases 3(a), 3(b), and 3(c).

Figure 3(c) also shows that even when (R3) holds, it is possible for inducement to succeed with a (p, q) detector but fail with (p', q) detector where $p' > p$. This means that inducibility is weak but not strong. This case occurs whenever the straight line separating the region where inducement succeeds (above) and the region where it fails (below) has a positive slope, as in Figure 3(c).

Recall that (R1) and (R2) imply that $a_{21} > a_{22}$ and, therefore, that the coefficient of q in (1) is positive. Differentiating (1) implicitly demonstrates that the slope of this straight line is $\frac{dq}{dp} = \frac{a_{12} - a_{11}}{a_{21} - a_{22}}$. It follows that the separating line has a nonpositive slope iff

$$a_{11} \geq a_{12}, \tag{R4}$$

which is a necessary condition for strong inducibility.

Because of (R1) and (R2), (R4) implies (R3). Thus, strong inducibility implies weak inducibility. In summary, weak inducibility occurs if (R1) – (R3) hold, and strong inducibility occurs if (R1) – (R4) hold. Strong inducibility occurs in cases 3(a) and 3(b), and weak inducibility in cases 3(a), 3(b), and 3(c).

In 2×2 strict ordinal games, Row can be weakly induced to choose *C* iff

$$a_{21} > a_{11} > a_{22} \text{ and } a_{21} > a_{12}$$

and strongly induced to choose *C* iff

$$a_{21} > a_{11} > a_{22} \text{ and } a_{11} > a_{12}.$$

In both cases, Row's maximum payoff must be a_{21}, which, in the absence of probabilistic tit-for-tat, would induce Row to depart from *CC*.

Finally, there are two cases of strong inducibility that are distinct when Column's detector has a characteristic curve $q = f(p)$. Assume Column chooses a value of p, say p_0, and then uses its detector with characteristic $(p_0, f(p_0))$, as discussed in section 2 and shown in Figure 2. By superimposing Figure 2 on Figure 3(a), it is clear that some characteristic curves may not intersect the (shaded) zone of inducement. Thus, in Figure 3(a), which occurs when $a_{21} > a_{11} > a_{22} > a_{12}$ (and is equivalent to $q_0 > 1$), it may be impossible to select detector characteristics so as to achieve inducement.

In this case, inducement is said to be *not guaranteed*. Of course, probabilistic tit-for-tat with a good enough detector, for which $f(p)$ is close enough to 1, will succeed, and inducement will be strong.

The situation is different for Figure 3(b), which occurs when $a_{21} > a_{11} > a_{12} > a_{22}$, In this case, it is clear that, by superimposing Figure 2 on Figure 3(b), it is always

possible to select some p_0 so that $(p_0, f(p_0))$ lies in the zone of inducement in the upper left of Figure 3(b). Note that Figure 3(b) corresponds to $q_1 \leq q_0 \leq 1$, for which inducement is said to be *guaranteed*. The games for which there is this guarantee are #7, #8, #9, #10, #13, and #14.

Figure 1

2×2 Game in Which Each Player Can Cooperate (*C*) or Defect (*D*)

		Column	
		C_C	D_C
Row	C_R	(a_{11}, b_{11})	(a_{12}, b_{12})
	D_R	(a_{21}, b_{21})	(a_{22}, b_{22})

Figure 2

Characteristic Curve for an Imperfect Detector

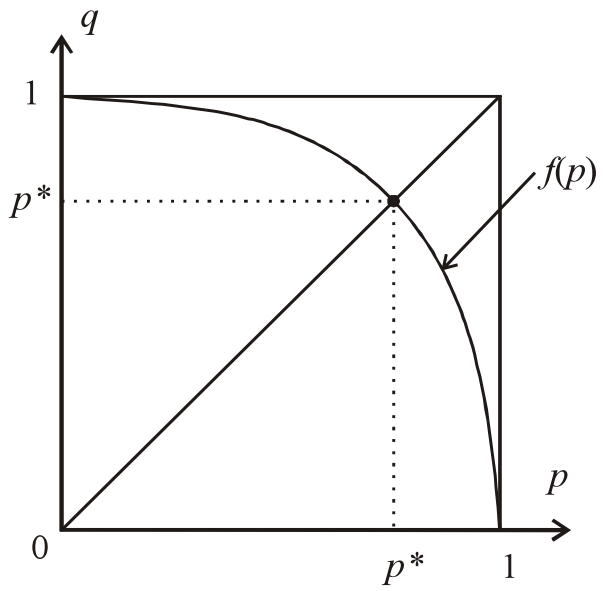
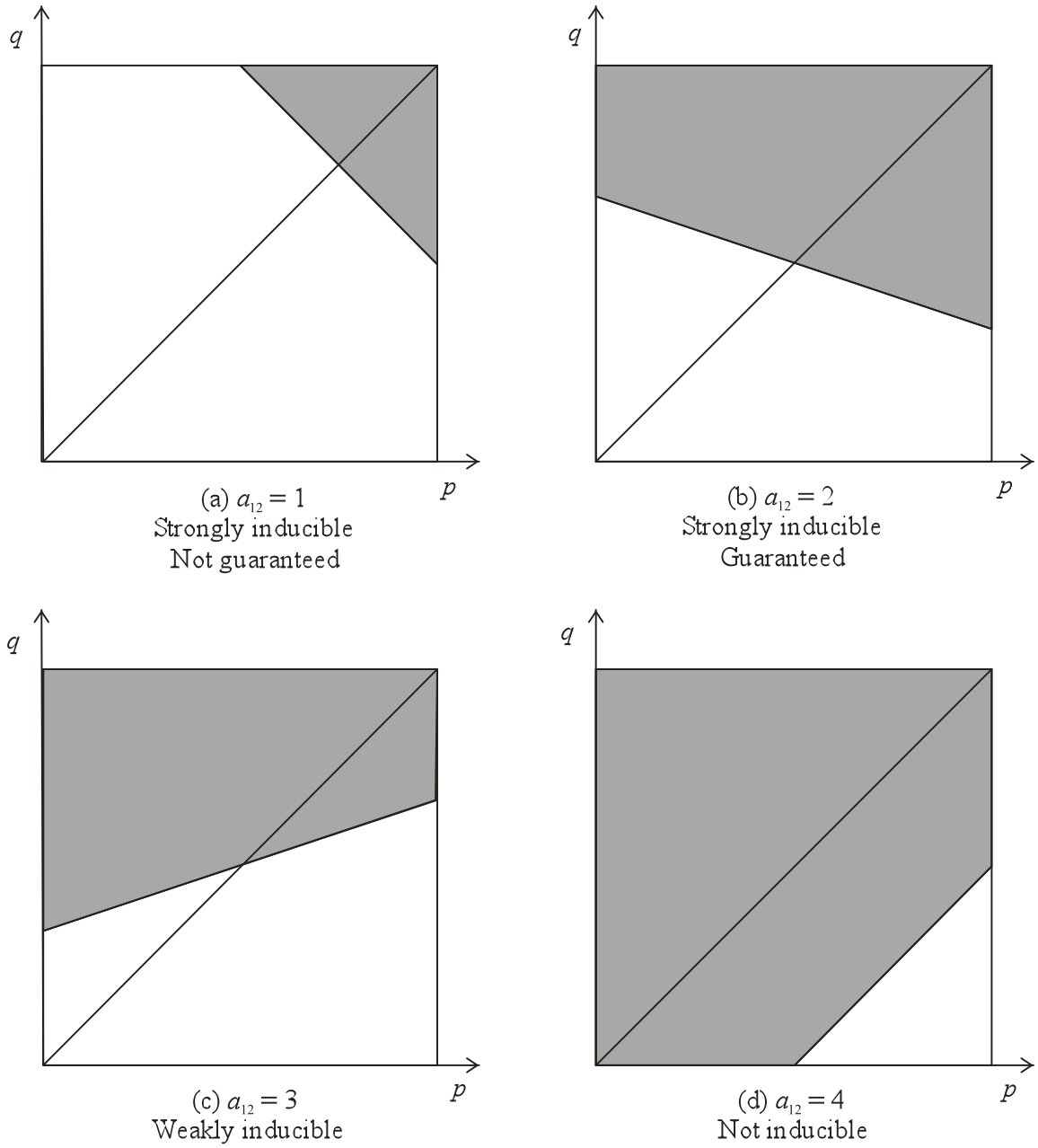


Figure 3

Region in (p, q) Unit Square where Inducement Can Occur (Four Cases)



Note

Weak and strong inducement are defined in both the text and the Appendix. Guaranteed and not guaranteed are defined in the Appendix.

Figure 4

Three Payoff Configurations for Column, and Eight for Row, of 24 Potentially Inducible Games

Column Configurations	Row Configurations		
	3, 2 4, 1	**3, 1 4, 2**	**2, 3 4, 1**
4, 3 2, 1	(3,4) (2,3) #50[d] (4,2) (1,1)	(3,4) (1,3) #35[d] (4,2) (1,2)	(2,4) (3,3) #56 (4,2) (1,1)
4, 3 1, 2	(3,4) (2,3) #29[d] (4,1) (1,2)	(3,4) (1,3) #28[d] (4,1) (2,2)	(2,4) (3,3) #47 (4,1) (1,2)
4, 2 3, 1	(3,4) (2,2) #37 (4,3) (1,1)	(3,4) (1,2) #33 (4,3) (2,1)	(2,4) (3,2) #39 (4,3) (1,1)
4, 2 1, 3	(3,4) (2,2) #31[d] (4,1) (1,3)	(3,4) (1,2) #27[d] (4,1) (2,3)	(2,4) (3,2) #45 (4,1) (1,3)
4, 1 3, 2	(3,4) (2,1) #36 (4,3) (1,2)	(3,4) (1,1) #34 (4,3) (2,2)	(2,4) (3,1) #38 (4,3) (1,2)
4, 1 2, 3	(3,4) (2,1) #46[d] (4,2) (1,3)	(3,4) (1,1) #48[d] (4,2) (2,3)	(2,4) (3,1) #43 (4,2) (1,3)
3, 4 2, 1	(3,3) (2,4) #57*[d] (4,2) (1,1) Ch	(3,3) (1,4) #22i (4,2) (2,1)	(2,3) (3,4) #53* (4,2) (1,1)
3, 4 1, 2	(3,3) (2,4) #22*[d] (4,1) (1,2)	(3,3) (1,4) #32[d] (4,1) (2,2) PD	(2,3) (3,4) #20* (4,1) (1,2)

Notes

1. Rankings of the payoffs to the players are as follows: 4 = best; 3 = next best; 2 = next worst; 1 = worst.
2. The numbers (#) of each game are those in the Appendices of Brams (1994, 2011).
3. Games #32 (Prisoners' Dilemma) and #57 (Chicken), the only symmetric games, are identified by PD and Ch, respectively.
4. Game #22i is obtained by interchanging Column and Row in game #22.
5. The inducible outcomes in all 24 games are those in the upper left, wherein Column is the inducer.
6. Pure-strategy Nash equilibria are underscored.
7. The four games with Nash equilibria that are better for Column than the inducible outcomes have an asterisk (*) after their numbers.
8. The 11 *difficult games* (see section 5) have a superscript 'd' after their numbers.
9. The eight games in the right-hand column are weakly but not strongly inducible; the remaining 16 games are weakly and strongly inducible.

Figure 5

20 Inducible Games and Column's Threshold Detection Probability (p_0)

<u>10 Difficult Games</u>

Class 1 (4 games)

1 (27)

(3,4)	(1,2)
(4,1)	<u>(2,3)</u>

$p_0 = ¾$

2 (28)

(3,4)	(1,3)
(4,1)	<u>(2,2)</u>

$p_0 = ¾$

3 (32)
Prisoners' Dilemma

(3,3)	(1,4)
(4,1)	<u>(2,2)</u>

$p_0 = ¾$

4 (48)

(3,4)	(1,1)
(4,2)	<u>(2,3)</u>

$p_0 = ¾$

Class 2 (3 games)

5 (22i)

(3,3)	(1,4)
<u>(4,2)</u>	(2,1)

$p_0 = ½$

6 (35)

(3,4)	(1,3)
<u>(4,2)</u>	(2,1)

$p_0 = ¾$

7 (50)

(3,4)	(2,3)
<u>(4,2)</u>	(1,1)

$p_0 = ½$

Class 3 (3 games)

8 (29)

(3,4)	(2,3)
(4,1)	(1,2)

$p_0 = ½$

9 (31)

(3,4)	(2,2)
(4,1)	(1,3)

$p_0 = ½$

10 (46)

(3,4)	(2,1)
(4,2)	(1,3)

$p_0 = ½$

Figure 5 (cont.)

20 Inducible Games and Column's Threshold Detection Probability (p_0)

10 Other Games

Class 4 (4 games)

11 (33)

(3,4)	(1,2)
(4,3)	(2,1)

$p_0 = ¾$

12 (34)

(3,4)	(1,1)
(4,3)	(2,2)

$p_0 = ¾$

13 (36)

(3,4)	(2,1)
(4,3)	(1,2)

$p_0 = ½$

14 (37)

(3,4)	(2,2)
(4,3)	(1,1)

$p_0 = ½$

Class 5 (6 games)

15 (38)

(2,4)	(3,1)
(4,3)	(1,2)

$p_0 = ½$ W

16 (39)

(2,4)	(3,2)
(4,3)	(1,1)

$p_0 = ½$ W

17 (43)

(2,4)	(3,1)
(4,2)	(1,3)

$p_0 = ½$ W

18 (45)

(2,4)	(3,2)
(4,1)	(1,3)

$p_0 = ½$ W

19 (47)

(2,4)	(3,3)
(4,1)	(1,2)

$p_0 = ½$ W

20 (56)

(2,4)	(3,3)
(4,2)	(1,1)

$p_0 = ½$ W

Notes

1. Rankings of the payoffs to the players are as follows: 4 = best; 3 = next best; 2 = next worst; 1 = worst.
2. The numbers (in parentheses of each game are those in the Appendices of Brams (1994, 2011).
3. Pure-strategy Nash equilibria are underscored.
4. Weakly inducible games are denoted by the letter 'W.'
5. Inducible outcomes are in boldface at the upper left.
6. For each game, the probability p_0 is the threshold value above which Column can induce Row to cooperate.

Figure 6

Iran-Israel Conflict: Two Games Wherein Iran Chooses to Develop (*D*) or Not Develop ($\overline{D}$) Nuclear Weapons and Israel Chooses to Attack (A) or Not Attack ($\overline{A}$)

Figure 6a: Game 1 (27)

		Israel	
		$\overline{A}$	A
Iran	$\overline{D}$	**(3,4)**	(1,2)
	D	(4,1)	(2,3)

$p_0 = ¾$

Figure 6b: Game 24

		Israel	
		$\overline{A}$	A
Iran	$\overline{D}$	(2,4)	(1,2)
	D	(4,1)	(3,3)

Notes

1. Rankings of the payoffs to the players are as follows: 4 = best; 3 = next best; 2 = next worst; 1 = worst.
2. Game 1 in Figure 6a refers to this game in Figure 5. The number that follows in parentheses in Figure 5a (27), and the number given for the game in Figure 6b (24), are those in the Appendices of Brams (1994, 2011).
3. Pure-strategy Nash equilibria are underscored.
4. The inducible outcome in game 1 in Figure 6a is shown in boldface in the upper left.
5. The probability (p_0) for game 1 (27) is the threshold value above which Israel can induce Iran to choose $\overline{D}$.

References

Axelrod, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.

Bergman, Ronen (2012). "Will Israel Attack Iran?" *New York Times Magazine*, January 30.

Biddle, W. F. (1972). *Weapons, Technology, and Arms Control*. New York: Praeger.

Biran, Dov, and Yair Tauman (2008). "The Role of Intelligence in Nuclear Deterrence." Preprint, Department of Economics, Tel Aviv University.

Brams, Steven J. (1994). *Theory of Moves*. Cambridge: Cambridge University Press.

Brams, Steven J. (2011). *Game Theory and the Humanities*. New York: MIT Press.

Brams, Steven J., Morton D. Davis, and Philip D. Straffin Jr. (1979a). "The Geometry of the Arms Race." *International Studies Quarterly* 23, no. 4 (December): 567-588.

Brams, Steven J., Morton D. Davis, and Philip D. Straffin Jr. (1979b). "A Reply to 'Detection and Disarmament.'" *International Studies Quarterly* 23, no. 4 (December): 599-600.

Brams, Steven J., and D. Marc Kilgour (1988). *Game Theory and National Security*. Oxford, UK: Basil Blackwell.

Brams, Steven J., and D. Marc Kilgour (1992). "Putting the Other Side 'On Notice' Can Induce Compliance in Arms Control." *Journal of Conflict Resolution* 36, no. 3 (September): 395-414.

Brams, Steven J., and D. Marc Kilgour (2009). "How Democracy Resolves Conflict in Difficult Games." In Simon A. Levin (ed.), *Games, Groups and the Global Good*. Berlin: Springer, pp. 229-241.

Bronner, Ethan (2012). "Israelis Assess Threats by Iran as Partly Bluff." *New York*

Times, January 27.

Bruns, Bryan (2011). "Visualizing the Topology of 2 × 2 Games: From Prisoner's Dilemma to Win-Win." Paper presented at the International Conference on Game Theory, Stony Brook, NY, July 11-15.

Dacey, Raymond (1979). "Detection and Disarmament: A Comment on 'The Geometry of the Arms Race.'" *International Studies Quarterly* 23, no. 4 (December): 589-598.

Erdbrink, Thomas (2012). "Iran Confirms Attack by Virus That Collects Information," *New York Times*, May 29.

Lindgren, Bernard (1976). *Statistical Theory*, 3rd ed. New York: Macmillan.

Molander, Per (1985). "The Optimal Level of Generosity in a Selfish, Uncertain Environment." *Journal of Conflict Resolution* 29, no. 4 (December): 511-518.

Nader, Aliteza (2012). "Influencing Iran's Decisions on the Nuclear Program." In Etel Solingen (ed.), *Sanctions, Statecraft and Nuclear Proliferation*. Cambridge: Cambridge University Press, 211-231.

Nowak, Martin A. (with Roger Highfield) (2011). *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. New York: Free Press.

Nowak, Martin A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, MA: Harvard University Press.

Robinson, David, and David Goforth (2005). *The Topology of 2 × 2 Games*. New York: Routledge.

Sigmund, Karl (2010). *The Calculus of Selfishness*. Princeton, NJ: Princeton University Press.

Solingen, Etel (2007). *Nuclear Logics: Contrasting Paths in East Asia and the Middle East*. Princeton, NJ: Princeton University Press.