ASYMMETRIC PROLIFERATION AND NUCLEAR WAR:
THE LIMITED USEFULNESS OF AN EXPERIMENTAL TEST

MICHAEL W. SIMON

International Relations Branch, SPARTA, Inc.
Arlington, Virginia, USA

I employ a human subject experiment to assess the relationship between nuclear proliferation and war. I develop a game-theoretic model to predict crisis behavior following four scenarios of dyadic nuclear acquisition. By varying the cash payments allotted to different outcomes, I experimentally alter the payoffs of the participants. Subjects compete for cash payments by playing a competitive game based on the model. The experimental variation (altering the cash payments) allows for an exploration of the difference between nuclear armament scenarios. Tentative results suggest that asymmetric nuclear acquisition would be dangerous. Yet, serious limits exist to the generalizability of experiments on nuclear proliferation, due to the gap between the laboratory setting and actual nuclear decision-making.

KEY WORDS: nuclear proliferation, asymmetric conflict, human subject experiments, international crisis behavior

How does the similarity of states’ capabilities affect the likelihood of war? This has long been an interesting question for international relations research, and is particularly important in considering the behavior of emerging nuclear rivals. The term “proliferation” implies that weapons spread to many countries. From a dyadic perspective,1 the timing of that process is important. What happens if rivals do not de-
ploy weapons at the same time? Is the likelihood of war different in that scenario from one in which proliferation is symmetric? I present a human subject experiment aimed at addressing these questions, based on predictions derived from a game-theoretic model. Subjects’ behavior differed from the game-theoretic predictions, leading to the tentative conclusion that asymmetric nuclear acquisition would be hazardous. There are important limits, however, to the applicability of experiments on nuclear proliferation, due to the gap between the impact of nuclear war and what can conceivably be simulated in a laboratory. Therefore, readers should be cautious about reading a great deal into the tentative conclusions. It is my intention for this article to add to the scientific debate over nuclear proliferation and war by presenting an experimental design, but I suggest that readers make their own conclusions about the usefulness of such an approach.

**LITERATURE**

*Traditional Approaches*

Realist theory argues that international politics is “anarchic,” and that countries must help themselves to achieve national goals (Waltz, 1979). Drawing from these realist precepts, Waltz (1995b) argues that proliferation would be beneficial. New nuclear states, according to Waltz, will deter each other, much as the United States and Soviet Union established Cold War deterrence by stockpiling nuclear weapons. Other writers use realist models to defend the nuclear activities of particular countries, notably India, Israel, and Pakistan (Chellaney, 1995; Feldman, 1982, 1995). In addition, the argument goes, nuclear capabilities of new proliferators will preclude fighting conventional wars, as well, since no country would be willing to risk nuclear devastation for the sake of winning a conventional battle (Waltz, 1995b, pp. 12, 16).

Using a bureaucratic modeling approach, Sagan (1995a) argues that proliferation would be hazardous. He agrees with Waltz and other realists that leaders will intend to be “rational” and safely control their nuclear weapons. Sagan believes, however, that there are limits to an individual’s ability to rationally pursue goals and objectives. The bureaucratic organizations that manage nuclear operations magnify these limits to rational behavior, leading to potentially catastrophic consequences. Thus, according to Sagan, it is inappropriate for researchers to use realist models; instead, political scientists should examine the internal workings of countries’ decision-making processes to determine the effects of proliferation.

How can this proliferation debate (Lavoy, 1995; Schneider, 1994) be resolved? Unfortunately for quantitative researchers (although perhaps fortunately for the rest of the world), very few countries have nuclear weapons. This general lack of proliferation makes it difficult to test empirically the claims made by Sagan, Waltz, and others about the effects of proliferation. Rigorous empirical tests are extremely rare (for exceptions, see Bueno de Mesquita and Riker, 1982; Simon, 1999). More common are loose discussions of a few cases (e.g., Ganguly, 1994; Karl, 1996/97; Waltz, 1995a. See Quester 1983 for an evaluation of this case-study literature). Others examine the superpower evidence and make conclusions about proliferation. (For broader studies on nuclear weapons possession and crisis behavior, see Geller, 1990; Harvey,

In a more general empirical study, Bueno de Mesquita and Riker (1982, pp. 289–292) examine three types of conflict—threats, interventions, and wars—from 1945 to 1976, and see how these conflicts are related to the possession of nuclear weapons. They conclude that deterrence seems to work—there are fewer conflicts involving nuclear countries, and those conflicts are less severe than the conflicts among nonnuclear countries. This research is interesting, but not definitive. It is not clear how Bueno de Mesquita and Riker approached the proliferation process—how a country moves from being nonnuclear to nuclear. In addition, their empirical research depends on a nonstrategic model of international behavior. They examine states’ decisions in isolation, rather than using a game-theoretic approach to include states’ anticipated reactions to their rivals’ future actions, and vice versa.

In a different set of empirical studies, Organski and Kugler (1980) and Kugler (1984) argue that nuclear deterrence was not a stabilizing force between the superpowers. As Kugler states, “[N]uclear weapons, despite their massive destructive potential, may not be the unique element that accounts for the absence of massive war in the international system in the last 35 years” (p. 501). Organski and Kugler argue instead that war and peace is determined by “power transition theory.” Believing that “the source of war is to be found in the differences in size and rates of growth of the members of the international system” (p. 20), Organski and Kugler conclude that “[t]he dominant nation and the challenger are very likely to wage war on one another whenever the challenger overtakes in power the dominant nation. It is this shift that destabilizes the system and begins the slide towards war” (p. 206). Kugler (1995), building on the result that “large disparities in power lead to peace” (p. 12), concludes that nuclear proliferation is dangerous only if both sides in a dyad have nuclear weapons (p. 24–25). (See also Kugler, 1998.)

Simon (1999) develops an incomplete-information game-theoretic model to link the choice to acquire nuclear weapons with the escalation of existing conflicts in a dyad. He focuses exclusively on asymmetric nuclear acquisition, and claims that asymmetric acquisition will be associated with a decrease in the severity of conflict. Simon’s model suggests two rationales for this explanation—deterrence and selection effects. According to the deterrence explanation, a nonnuclear state will back down in a crisis when facing a nuclear state, fearing unacceptable damage. According to the selection effects explanation, a state with low expected utility for fighting a conventional war will acquire nuclear weapons, while a state with a high expected utility for fighting a conventional war will remain nonnuclear. In other words, if a state does not fear a conventional battle (because it has a favorable conventional balance of forces, cares about the issues at stake, is willing to take chances, etc.), then it will not pay the costs of acquiring nuclear weapons. Simon empirically tests the converse of this hypothesis, focusing on the impact of choosing not to unilaterally acquire nuclear weapons. He finds that this choice is a significant predictor of conflict severity, while controlling for other factors related to conflict.

Bueno de Mesquita examines the history of poisonous gases used in warfare, particularly in World War I and World War II. He claims that the significant use by Germany of poison gas during World War I “was encouraged in part by the allied failure to develop a retaliatory capability that could impose like costs on the Ger-
mans” (Bueno de Mesquita, 1988, p. 10). In World War II, however, chemical weapons were only used against enemies that could not respond with their own chemical weapons (pp. 12–14). Bueno de Mesquita notes further that this pattern continued after World War II, with the incidents of chemical attack (by the U.S. in Vietnam, by the USSR in Afghanistan, by the USSR and/or Vietnam in Kampuchea, and by Iraq against Iran) occurring only when the recipient did not possess the capability to “respond in kind” (p. 19).2 Bueno de Mesquita argues that behavior with nuclear weapons is likely to be similar to behavior with chemical weapons. Therefore, he suggests that the U.S. promote nuclear proliferation, in selective cases, in order to foster symmetric deterrence.

The recent research examines history from a variety of methodological perspectives, but appears unable to provide a conclusive prediction of the effects of proliferation. Among political scientists, debates today over the results of proliferation are as strong as they have ever been. Some observers claim that these debates continue, in part, due to inadequate examination of the available evidence (Feaver, 1995). Although I agree with Feaver in principle on this point, I feel that a proper empirical study first requires better theory. In addition, researchers may find alternate data sources to be useful. The next section describes some attempts to simulate nuclear acquisition and assess proliferation arguments in a hypothetical environment.

Simulation and Modeling Approaches

Some researchers, evidently concerned with proliferation but dissatisfied with the usefulness of empirical evidence on the subject, employ techniques to study the effects of nuclear proliferation that do not depend on historical data. Those techniques include human simulations and experiments, and formal modeling. While these approaches are innovative and interesting, existing applications generally fail to provide conclusive predictions about the effects of proliferation. I describe the simulations literature first.

Brody (1963) adapts an existing role-playing model, the Inter-Nation Simulation (I.N.S.), to simulate the international environment before and after nuclear proliferation.3 Brody’s subjects were Chicago-area high school students, chosen for their ability to think in abstract terms. He assigned each student two of the traditional I.N.S. leadership roles so that (in his simulation) a country’s “cabinet” operated as follows: One student played the roles of the Central Decision Maker and the Interior Decision Maker; one student played the roles of External Decision Maker #1 and the Decision Maker for Force; the other student played the roles of External Decision Maker #2 and the Representative to the International Organization. During the simulation, each cabinet received updates on how satisfied the (theoretical) domestic population was in terms of economic security and national security. If the level of domestic satisfaction fell too far, the cabinet would be “overthrown.”

In each of the twelve half-hour rounds, each country had to fill out a “decision sheet” to declare what action it would take during the next round. Some of the possible actions included fight a large or small war, trade, aid, make alliance, or communicate (by letter, face-to-face, in the international organization, or through the world press). In the first few rounds, the “world” was divided into two blocs; leading each
broc was a “country” with nuclear weapons. (So at the beginning of the simulation, only two countries possessed nuclear weapons.) In the middle of the simulation, all of the countries in the world acquired nuclear weapons.

Brody’s aim was to see if the spread of nuclear weapons technology influenced bloc behavior. Using content analysis, Brody examined international communications before and after the “experimental intervention” (the universal acquisition of nuclear weapons). While the simulated countries moved in the direction Brody had expected, they did not go as far. He had hypothesized that proliferation would lead to the breakup of pre-existing bloc alliances. According to the results of the content analysis, blocs were weakened but not destroyed as a result of proliferation.

While Brody’s experimental intervention attempts to simulate the proliferation of nuclear weapons, his method glosses over an important aspect of the proliferation process: Will new nuclear states (i.e., ones with minimal nuclear forces) be able to achieve deterrence (Sagan, 1995b, pp. 126–128; Waltz, 1995c, pp. 108–110)? Standard deterrence stories depend on each side having enough weapons to be able to survive an attack and destroy the original attacker (Berkowitz, 1985; Bueno de Mesquita and Riker, 1982; Intriligator and Brito, 1981; James, 2000b). In the first few rounds of Brody’s simulation, only two of the countries (analogous to the U.S. and the USSR) had nuclear weapons. After the experimental intervention, all of the countries in the simulated world had nuclear weapons. However, nothing distinguished the nuclear capability of the newly acquired states’ weapons from the weapons of the two bloc-leaders. Writing in 1969, de Leon et al. note: “Since nuclear countries like Britain, France, and China have not been able to acquire capacities even roughly commensurate to those of the United States and the Soviet Union, the dissemination of equivalent nuclear weapons technology is a distinctly unrealistic presumption of [Brody’s] experimental intervention” (p. 53). Beyond some magnitude, is the degree of disparity relevant? This has long been a question for nuclear analysts (Nalebuff, 1988; Rosenberg, 1984; Schilling, 1984). Brody (1963) answers the question by implicit assumption, rather than allowing his simulation to reach a conclusion. Setting aside the overall validity or usefulness of Brody’s simulation, his design precludes a direct assessment of the effects of asymmetric nuclear proliferation.

Other researchers use formal models to simulate various nuclear environments. This technique is beneficial in that it forces researchers to clearly state their assumptions and arguments (eliminating a major problem with the traditional literature) and also in that it allows researchers to make theoretical arguments in the absence of extensive amounts of data. Unfortunately, the results from much of this research to date generally depend on implicit assumptions about the distribution of capabilities resulting from proliferation.

In particular, the formal literature from the 1980s mis-specifies a very key concept, war. Formal work from that period (Intriligator and Brito, 1981; Bueno de Mesquita and Riker, 1982; Berkowitz, 1985) is generally supportive of proliferation. This support may be due to the use of restrictive definitions. Intriligator and Brito’s overall conclusion is that nuclear proliferation would decrease the likelihood of war because new nuclear states would deter each other. Their model, however, implicitly defines “war” as what happens when two nuclear states attack each other (p. 258). Thus, one of the outcomes feared by Sagan and others (e.g., Dunn, 1982, 1991), a
nuclear state attacking a nonnuclear state, is not allowed for in Intriligator and Brito’s model.

Bueno de Mesquita and Riker (1982) and Berkowitz (1985) also conclude that nuclear proliferation can be beneficial. However, these conclusions are drawn, in part, from the restrictive assumptions of their respective models, each of which eliminate some of Dunn and Sagan’s concerns. Bueno de Mesquita and Riker assume (in the first of two models) that nuclear states will not attack each other (pp. 287–288). Berkowitz assumes that states possessing a second-strike capability will not attack each other (pp. 121–122). This is not to say that building models based on restrictive assumptions is wrong; I very strongly endorse the use of restrictive assumptions. However, readers must be aware of those assumptions when examining broad conclusions such as proliferation decreases the likelihood of “war.”

Bueno de Mesquita and Riker (1982) begin to make these needed distinctions in their second model (pp. 292–303). Using an expected-utility framework, they model the decisions one state (the initiator) would make when facing another state (the defender) in a crisis. Their model covers four different situations:

(A) The initiator and the initiator’s opponent both have a nuclear capability sufficient to impose unacceptable damage on the other.
(B) The initiator has a nuclear capability as described in (A), but the opponent has only a conventional (or very modest nuclear) capability insufficient to impose unacceptable damage.
(C) The initiator has only a conventional capability and the opponent has a nuclear retaliatory capability as in (A).
(D) Both the initiator and the defender have only a conventional (or very modest nuclear) capability insufficient to impose unacceptable damage (Bueno de Mesquita and Riker 1982, p. 293).

Given an ordering of preferences for the initiator, Bueno de Mesquita and Riker calculate the expected utility gained and lost for fighting under conditions A through D. They conclude that, under most conditions, the initiator prefers (B) to (D) to (A) to (C) (p. 299); in (B), the initiator can most convincingly threaten the defender, since only the initiator has nuclear weapons. In (C), only the defender has nuclear weapons, and so the initiator prefers this situation the least.

Bueno de Mesquita and Riker thus present the case for “selective proliferation” (p. 303). They argue that, if “proliferation” means only one state in a rivalry acquires nuclear weapons (akin to (B) or (C)) that state would have a great advantage over its rival. However, if “proliferation” means that a rivalry moves from a situation in which only one state has nuclear weapons to a situation when both have nuclear weapons, the prospects for peace are greatly enhanced since neither country can threaten the other.5

Of the formal models discussed so far in this section, Bueno de Mesquita and Riker’s (1982) second model most closely includes the essential elements of proliferation. This model allows for an examination of different conflict scenarios involving nuclear weapons. However, the situation in which a country possesses only a minimal nuclear capability—a snapshot in time immediately after a country has “pro-
liferated”—is only allowed for in the verbal explanation of their model (p. 293) and not formally expressed in their mathematical model (p. 295). In order to decisively test the impact of nuclear acquisition, a formal model should incorporate this scenario. Thus, some key elements of proliferation still lack formal attention.

In addition, only one player’s decisions are mapped in Bueno de Mesquita and Riker (1982). Instead of a decision-theoretic model, researchers would find a game-theoretic approach more useful, since models of the latter type can be used to examine the strategic interaction between two (or more) players as they react to the actions of each other. Previous research in international conflict suggests the usefulness of game-theoretic models at describing strategic behavior (e.g., Bueno de Mesquita and Lalman, 1992; Fearon, 1994a; Fearon, 1994b; Fearon, 1997; Smith, 1995).

Three recent studies present game-theoretic models that address some key proliferation situations. Simon (1999, see previous section) constructs an asymmetric-information game-theoretic model that focuses on selection bias. His model, while empirically tested, only considers the impact of one nuclear state in a dyad. Simon concludes that asymmetric acquisition would enhance peace, but contrasts that situation with nonproliferation (rather than symmetric acquisition). Simon suggests that leaders decide to acquire nuclear weapons based on their expected utility for fighting a conventional war, and that the decision not to acquire weapons communicates important information to other players. This signaling opportunity is beyond the scope of the present article.

James (2000b) classifies nuclear arsenals into four categories, depending on each state’s warhead capability and destructive capacity. In developing a game-theoretic model based on the Theory of Moves, James assigns players to a particular preference ordering depending on their arsenal category. She then explores one of the implications of this model with a discussion of Iran and Iraq, two states with mini-arsenal capacity (see also James, 2000a). It remains further to address the additional implications of the model.

Kraig (1999) presents a complete-information model that distinguishes between nuclear and conventional responses to crises. Kraig’s model allows for many different preference orderings of the players, represented by ordinal rankings of the outcomes in the game. The key for Kraig is whether players can make credible threats to use either conventional or nuclear weapons. His model produces results for 14 scenarios of dyadic nuclear development (p. 158); mutual cooperation is predicted in four cases, one-sided cooperation and one-sided defection is predicted in eight cases, and mutual defection is predicted in two cases. Kraig concludes overall that nuclear proliferation makes stability difficult to achieve. Although illustrated with historical examples, Kraig’s predictions have not been subjected to statistical verification.

In the next section, I discuss an experiment I performed that was designed to study the impact of asymmetric proliferation on the likelihood of war. Included in the experimental design is a simple game-theoretic model, designed to allow for variation in the symmetry of nuclear capabilities. While the model is simpler than in Kraig (1999), the essential characteristics are the same. Additionally, the use of a simple model allowed for an experimental assessment of whether decision-makers actually behave according to the game-theoretic predictions.
EXPERIMENTAL DESIGN

Although not often used in studying international relations, experiments are a rich and important tool in many areas of political science (Kinder and Palfrey, 1993; Palfrey, 1991). The main advantage to using experiments in studying politics is no different from the rationale in other scientific fields—a laboratory allows a scientist to control various factors and isolate the phenomena of interest. Role-playing simulations and human experiments share many characteristics, most notably that the subjects behave in an artificial environment designed by the researcher. In a role-playing simulation, the subjects know that they are part of a simulation, and are asked to pretend that they are other people. In an experiment, similarly, the subjects know that they are involved in a laboratory setting. Experimental subjects, however, are asked only to respond naturally to their (artificial) surroundings. Thus the basic appeal of experiments over simulations is that the results do not depend on the subjects’ ability to role-play and pretend. Mintz et al. (1997), for example, instruct their simulation subjects that the “quality of the decision you make in the context of the simulation will suggest your ability to comprehend national level decision making” (p. 559). If an experiment is well designed, the subjects can behave without feeling constrained to a particular role, and yet the experimenter can extrapolate from this limited environment to some real-world situation.

In three separate days, I ran subjects through an experiment based on a game-theoretic model of asymmetric nuclear acquisition. The purpose of the experiment was to see whether the subjects would behave according to the model’s predictions regarding international conflict. The subjects recruited for this experiment were all undergraduate students; using undergraduates in academic experiments is quite common. I told potential subjects that they would be participating in a “decision-making experiment” that would require them to sit at desks and fill out materials. I also told them that they would be paid in cash at the end of the experiment. If they pressed for more information, I said: “You will earn approximately fifteen dollars, but the exact amount you will earn will depend on the decisions you make in the experiment.” As will be discussed in a later section, I employed two experimental treatments to control for differences between expected benefits and expected losses.

Thirty-two subjects were used on three separate days—ten on each of the first two days, and twelve on the last day. Each day’s experiment session lasted approximately two hours and consisted of ten game rounds. The subjects were split into pairs and each member alternately marked their move on a game tree diagram. This continued until the subjects reached the end of the round. At the beginning of the session, each student received a copy of the instructions, an earnings sheet, and two sample diagrams. The instructions were read aloud by the experimenter. I can best summarize the experiment by listing the instructions that the subjects received:

Instructions

This is an experiment testing your decision-making ability. You will earn money, which we will pay to you in cash at the end of the experiment. The amount of money you will earn depends on the decisions you will make.

Please examine the page labeled “Sample Diagram One.” This page illustrates
the experiment we will be doing. During different rounds of the experiment, you will be either player A or player B. You and a randomly chosen opponent will alternate deciding whether to move left or right. The game continues until you reach one of the sets of numbers. In the sample diagram, if player A moves left, the game stops and both players earn $1.50 for that round. Suppose instead that player A begins by moving right, and then player B moves right, and then player A moves left. The round is over and player A earns $1.46 and player B earns $1.50.

We will begin each round by handing out folders to half of you, who will be player A for that round. You will record your move on the diagrams by circling “left” or “right.” When you are finished recording your move, close your folder. We will then pick up the folders and show your move to your opponent, who will circle left or right. This will continue until the round ends. At the end of each round, you can record how much money you made by writing it on your earnings sheet. For example, if you as player A earn $1.50 in round one, write that down in the spaces provided on your earnings sheet.

You will be using different diagrams throughout the experiment. Some of the diagrams look like Sample Diagram Two. Please look at it now. If you are player A, and you move right, and then player B moves right, and then you move right again, you will pick a token to determine how much money you will earn.

This bucket contains twenty tokens, labeled one to twenty. During the experiment, if your move requires you to pick a token, raise your hand. We will come by with the token bucket and you will randomly pick one of the tokens. Look at the number on the token, and write down that number in the space provided on the diagram. On Sample Diagram Two, for instance, if you picked token number twenty, you (as player A) would write down “20” in the space marked “token number.”

Figure 1. Sample Diagram One (same diagram as Scenario I).
would receive $1.40, and player B would receive $1.50.

We will keep a record of your earnings on the official list, but you should keep your own list as well on the Earnings Sheet.

No talking will be allowed during the experiment, which will begin in just a moment. If you talk during the experiment, you will be asked to leave and you will forfeit your earnings. There will be ten rounds of this experiment. A reminder: When you have finished circling your move, please close your folder. The experiment will now begin.

A “round” consisted of the subjects moving through a game tree diagram, which resembled the game displayed in Figure 4. The game proceeds sequentially. Player A either makes a demand against Player B, or the game ends with the status quo outcome. If a demand is made, Player B responds by giving in to the demand or not giving in. If B does not give in, A moves again by attacking B or backing down. This game is similar to the standard deterrence model, but differs in specification from the most well known version in the literature. In order to model nuclear war, the concept of war used here differs from the oft-cited model of Bueno de Mesquita and Lalman (1992). In their International Interaction Game (p. 30), war between two states can

Figure 2. Sample Diagram Two (same diagram as Scenario II).
only occur if both states decide to fight each other. While this is a useful concept when discussing conventional war, I eliminate the bottom stages of the International Interaction Game in order to focus on nuclear attack only. Since I am interested in modeling cases in which one, both, or none of the states in a dyad have nuclear weapons, the Nuclear War Game employs the following definition of nuclear war—when one state attacks another with nuclear weapons. This definition is also consis-

---

**EARNINGS SHEET FOR**

(please indicate your name)

(Social Security number)

Round | Which player was I? | Amount I earned
---|---|---
1 | | .
2 | | .
3 | | .
4 | | .
5 | | .
6 | | .
7 | | .
8 | | .
9 | | .
10 | | .

Total_____.

Please fill in your seat number here: ____

---

Figure 3. Earnings Sheet.
tent with the only historical use of nuclear weapons in wartime—the U.S. bombing of Japan.

I took several precautions to ensure that the subjects would only communicate with each other in a manner controlled by the experiment. First, the subjects were not allowed to talk during the experiment, aside from asking the experimenter questions if they were confused. The subjects were told that they would be forced to leave the experiment without payment if they talked to each other. Aside from asking clarification questions, none of the subjects spoke during the experiment. Second, the subjects sat at desks in a semicircle in a classroom. After listening to the instructions, the subjects were told to turn their desks away from the center of the room, so that each subject performed the experiment while facing the wall.

Last, I ran the experiment in a manner making it very difficult for the subjects to know the identity of their opponent each round. The materials handed out to and picked up from each student were enclosed in manila folders, which all looked identical. Hidden markings on the back of the folders identified them to assist in running the experiment. Additionally, all the folders were collected before any were handed out to the next player.

These precautions were employed in order to be able to treat each round-pair as an independent observation. Davis and Holt (1993, p. 528) note the importance of such precautions in order to reduce the possibility of future periods being influenced by one “unusual and bizarre” subject’s behavior.

![Figure 4. Nuclear War Game.](image)
Four game tree diagrams were used in the experiment, each resembling Figure 4 (the Nuclear War Game). The diagrams represented four different scenarios of proliferation. In Scenario I (Fig. 1, the same as in the instructions) neither Player A nor Player B possesses nuclear weapons. In Scenario II (Fig. 2, also as in the instructions) Player A possesses a few nuclear weapons and Player B possesses none. In Scenario III (Fig. 5) both players possess a few nuclear weapons. In Scenario IV (Figure 6) both players possess many nuclear weapons. The payoffs for each outcome differ depending on the scenario. The “nuclear war” outcome in Scenario III, for example, is not as severe as that same outcome for Scenario IV.

Clearly, the dollar amounts of the payoffs are a far cry from the costs and benefits of actual international decisions. Instead, the payoffs are part of an internally consistent formal model of nuclear crisis bargaining, wherein players are assumed to maximize expected utility. For obvious practical reasons, it would be impossible to deliver payoffs to experimental subjects that amount to the real impact of nuclear war. Instead, the monetary amounts for the payoffs were chosen to reflect the relative

---

**Figure 5.** Scenario III.

---

---
amount of damage that a nuclear exchange would bring, when comparatively assessed between scenarios. These payoffs were then scaled to a level consistent with what could be paid out in an experimental setting. It is reasonable to state that the payoff structure is about as close as one could come to mimicking nuclear war in an experimental situation.

The payoffs for the bottom right node of Scenarios II, III, and IV were determined by picking a random token. This design feature was used to try to incorporate the “uncertainty” leaders would have about the outcome of a nuclear war. Even in situations when leaders know how many nuclear weapons their states and enemy states have and have a fairly good idea how much damage would result from a war, part of the mystique of nuclear weapons is due to the fact that no one really knows exactly what would happen if they were used. This uncertainty may, in fact, be advantageous. Waltz notes: “States are not deterred because they expect to suffer a certain amount of damage [in a nuclear war] but because they cannot know how much damage they will suffer.” (1995b, p. 34, emphasis added. See also Waltz, 1995b, pp. 6,
All subjects in the experiment thus only had a probabilistic estimate of what the payoffs would be for the bottom right node of the game. The actual payoffs were determined when the player picked a token, which occurred after the decision to move down the right side of the tree.

Although Figures 1, 2, 5, and 6 represent nuclear conflict, this fact was not revealed to the subjects. Instead, subjects were told that they would move left or right and were playing a “decision-making” game. At the beginning of each round, half of the subjects were handed a folder with a game tree diagram. The distribution of diagrams was determined by a random drawing performed well before the experiment. Subjects receiving folders were designated as “Player A” for that round. When the A players finished moving left or right they closed their folders and the folders were distributed to their designated opponents for that round (B players). Which player was A or B and who played who was also determined by a random drawing before the experiment.

A spreadsheet was examined during the experiment in order to keep track of how to pass out the folders. Figure 7 is a section of the first day’s spreadsheet, indicating the setup of round 3. Players 8 and 10 played each other using diagram II, 5 and 9 played each other using diagram III, and so on.
Cash Payments

The use of cash payments is increasingly employed in experiments in political science and in economics in order to induce preferences in subjects (e.g., Gerber et al., 1998; Wilson and Rhodes, 1997; see also Kinder and Palfrey, 1993; Palfrey, 1991, pp. 1–3). Previous experiments have established the practical benefits of cash payments, which range from a reduction in performance variability (Davis and Holt, 1993, pp. 24–25) to increased gains in external validity, when compared to role-playing simulations (Davis and Holt, 1993, p. 450; Morgan and Wilson, 1989, p. 2). While we should be skeptical of extrapolating laboratory results to the real world, we should be much less suspect of this type of experiment than of a role-playing simulation testing the effects of proliferation. Subjects in role-playing simulations have no incentive to behave “truthfully”; in a cash payment experiment, however, subjects have a financial incentive.9

Treatments

The first two days’ subjects were given the initial treatment. On those days, the subjects were paid in cash at the end of the experiment. They did not see any cash until the experiment concluded. On the last day’s experiment, the subjects were each handed an envelope with $15 in cash at the beginning of the session. The subjects in this treatment received the same instructions as in the other treatment, except the first paragraph read as follows:

This is an experiment testing your decision-making ability. We are now handing out $15 in cash to each of you. During the experiment, you will earn money. The amount of money you will earn depends on the decisions you will make. If, at the end of the experiment, you have earned more than $15, we will give you the extra money you have earned. If you have earned less than $15 during the course of the experiment, we will take the difference out of that $15.

The envelopes were collected after the instructions were read, and were placed on a chair in the middle of the room. The two treatments were employed to avoid forcing a bias in the experiment toward one risk preference or another. Intuitively, people may react differently to gains and to losses, as the following example from Prospect Theory indicates: Suppose you were given a choice of playing two lotteries. In lottery A, you have a 10 percent chance of winning $9, and in lottery B you have a 90 percent chance of winning $1. Which would you choose? Suppose now you are given a choice of playing two different lotteries. In lottery C, you have a 10 percent chance of losing $9, and in lottery D you have a 90 percent chance of losing $1. Which lottery would you choose?

Many people tend to feel differently about potential gains and potential losses, and would choose lottery A in example 1 and lottery D in example 2 (despite the fact that the Expected Value of the gain or loss in all four lotteries is exactly the same: .9).

Note that there are no negative payoffs for reaching any outcome in the experiment, in the sense that the payoffs are all greater than or equal to zero. This is for simplicity’s sake—the worst outcome is normalized to zero. What is important is
that the “bad” consequences of war all result in a loss of money compared with what the status quo would have produced. In Scenario IV, for example, if A moves left, it receives $1.50. If it reaches the bottom right outcome, and token 1–19 is chosen, then A receives $0. An alternative approach would have required the experimenter to confiscate money that the students themselves brought to the laboratory. In designing the experiment, I was not able to find a way to achieve this (ostensibly) desirable mechanism while maintaining scientific control and still follow rules ordained by the university’s Human Subject Committee.

HYPOTHESES

Backwards-induction outcomes were calculated for each scenario of the game. Assuming that all players were risk-neutral profit-maximizers, the players were expected to play in order to reach the maximum personal payoff. In Scenario II (Fig. 2), for instance, Player A’s expected value for picking left at the bottom stage was $1.20, and $2.57 for picking right. Realizing that Player A would thus choose to go right in...

| Table 1 |

<table>
<thead>
<tr>
<th>Raw Experimental Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
</tr>
<tr>
<td>Scenario I</td>
</tr>
<tr>
<td>AL*</td>
</tr>
<tr>
<td>AR-BR-AL</td>
</tr>
<tr>
<td>AR-BL</td>
</tr>
<tr>
<td>Scenario II</td>
</tr>
<tr>
<td>AR-BL*</td>
</tr>
<tr>
<td>AL</td>
</tr>
<tr>
<td>AR-BR-AR</td>
</tr>
<tr>
<td>AR-BR-AL</td>
</tr>
<tr>
<td>Scenario III</td>
</tr>
<tr>
<td>AL*</td>
</tr>
<tr>
<td>AR-BL</td>
</tr>
<tr>
<td>AR-BR-AR</td>
</tr>
<tr>
<td>AR-BR-AL</td>
</tr>
<tr>
<td>Scenario IV</td>
</tr>
<tr>
<td>AL*</td>
</tr>
<tr>
<td>AR-BR-AR</td>
</tr>
<tr>
<td>AR-BR-AL</td>
</tr>
</tbody>
</table>

Notes:

- “*” indicates the predictions of the model
- “R” indicates a player moved right
- “L” indicates a player moved left
- Example: “AR-BL*” means that A moved right and B moved left, and that this was predicted by the model.
the last stage of the game, Player B would compare a payoff of $.50 for moving left in stage two against the eventual expected value $.42. Realizing that Player A would move right in the last stage if it was given the chance (since 2.57 > 2.5) Player B would move left in the second stage. At the start of the game, Player A would thus be anticipating $2.50 if it moved right in the first stage and $1.50 if it moved left. The backward-induction outcome (and the game-theoretic prediction) for this scenario was A-Right, B-Left. By assumption, the value for money is set by outside forces (the market), and is not determined by how other players value money. In that sense, there is no monetary winner or loser.

The predicted outcomes for the other scenarios are as follows: Scenario I: Either A-Left ( = Status quo), or A-Right, B-Right, A-Right ( = War). Scenarios III and IV: A-Left ( = Status quo). Recall that for Scenario I, neither player has nuclear weapons. Since neither player can launch weapons, the outcome of the bottom right node is the same as the status quo. In Scenario II, the model predicts that Player A will threaten and Player B will give in. In Scenarios III and IV the model predicts that Player A will not threaten. Note that the model never predicts nuclear war.

RESULTS

The raw data are listed in Table 1. When presented with Scenario I (neither side with nuclear weapons), subjects overwhelmingly behaved according to the model’s predictions (95 percent of the time). In the other scenarios, however, the subjects often did not behave in the predicted manner. For Scenario II (Player A with a few nuclear weapons, Player B with none), the model was accurate only 19 percent of the time; for Scenario III (both players with a few nuclear weapons), 62 percent; for Scenario IV (both players with many nuclear weapons), 77 percent.

Following Davis and Holt (1993, pp. 533, 546), the \( \chi^2 \) test was used to compare the two treatments (money given after the experiment vs. money given before the experiment). For each scenario, the \( \chi^2 \) test confirms that there was no difference in the behavior of subjects from one treatment to the other (see Table 2).

If the subjects behaved according to the predictions from the game-theoretic model, the results from that round are recorded as “yes.” “no” responses indicate that the subjects’ behavior was inconsistent with the model’s predictions. The numbers in

---

### Table 2

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$$$ After</strong></td>
<td>24</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>23.7</td>
<td></td>
<td>1.3</td>
</tr>
<tr>
<td><strong>$$$ Before</strong></td>
<td>12</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>12.3</td>
<td></td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>36</td>
<td>2</td>
<td>38</td>
</tr>
</tbody>
</table>

Note: Observed data listed in regular type; expected data listed in italics.
regular print are the observed data, and the numbers in italicized print are the expected data.

The expected data were calculated as follows: Upper left: \((25/38) \times (36/38) \times 38 = 23.7\). Upper right: \((25/38) \times (2/38) \times 38 = 1.3\). Bottom left: \((13/38) \times (36/38) \times 38 = 12.3\). Bottom right: \((13/38) \times (2/38) \times 38 = 0.7\).

Since we are interested in whether the two treatments are different from one another (and not simply whether they are different in only one direction), it is appropriate to use a two-tailed test. The degrees of freedom = \([(\# \ of \ rows - 1) \times (\# \ of \ columns - 1)] = 1\). In order to reject the null hypothesis of no difference between the two treatments (at 95 percent confidence), we would have to have \(\chi^2 \geq 3.84\). For Scenario I: \(\chi^2 = [(24 - 23.7)^2 / 23.7] + [(1-1.3)^2 / 1.3] + [(12 - 12.3)^2 / 12.3] + [(1 - 0.7)^2 / 0.7] = 0.2089\). Thus for Scenario I the null hypothesis of no difference between the treatments cannot be rejected. The same procedure was followed to test the other scenarios. The \(\chi^2\) value for Scenario II is 1.09; the \(\chi^2\) value for Scenario III is 0.18; and the \(\chi^2\) value for Scenario IV is 0.37. Thus, the null hypothesis cannot be rejected for any of the scenarios. This analysis indicates that there was no difference in the behavior of subjects from one treatment to the other in any scenario. The data for both treatments is pooled in the analysis below.

The \(\chi^2\) test was used to compare the results of each scenario. In five out of six cases, the \(\chi^2\) value exceeded the critical value (see Table 3); therefore, the null hypothesis of no difference between scenarios must be rejected in those cases. The exception was the case of comparing Scenarios III and IV, in which the null hypothesis cannot be rejected.

Most, but not all, of the deviations from the model’s predictions occurred when the subjects reached the nuclear war outcome. The following comparisons were made after excluding from the data those cases in which subjects did not behave as pre-

### Table 3
Testing the Accuracy of the Risk-Neutral Predictions

<table>
<thead>
<tr>
<th>SCENARIO</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>43.06*</td>
<td>12.33*</td>
<td>4.99*</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>13.68*</td>
<td>24.74*</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td></td>
<td>2.17</td>
</tr>
</tbody>
</table>

* = exceeds the \(\chi^2\) critical value, must reject the null hypothesis of no difference between scenarios.

### Table 4
Data Comparison on Nuclear War

<table>
<thead>
<tr>
<th>SCENARIO</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>12.33*</td>
<td>27.50*</td>
</tr>
<tr>
<td>III</td>
<td>4.61*</td>
<td></td>
</tr>
</tbody>
</table>

* = exceeds the \(\chi^2\) critical value, must reject the null hypothesis of no difference between scenarios.
dicted, but also did not reach the nuclear war outcome. The number of cases ex-
cluded were 6 for Scenario II, 6 for Scenario III, and 5 for Scenario IV.

The \( \chi^2 \) test was used to compare the results of each scenario. In all three cases, the \( \chi^2 \) value exceeded the critical value (see Table 4); therefore, the null hypothesis of no difference between scenarios must be rejected in those cases.

**CONCLUSIONS**

*Game-theoretic Predictions and Experimental Results*

While it may seem surprising\(^{12}\) that the game-theoretic model predicted so poorly, these results do not at all discredit the use of formal models to analyze proliferation and war. The precision of the Nuclear Weapons Game made it relatively easy to design a test of the model, and even though the model’s predictions were often not borne out experimentally, the model served as a useful benchmark for analyzing the experimental results.

*Proliferation and Nuclear War*

The ambiguity of the results of this experiment does not shed much light on the debate between Sagan and Waltz over how easy it will be for new nuclear states to establish deterrence. Sagan argues that states will have to acquire many weapons for deterrence to exist, while Waltz feels that only a minimal nuclear force is necessary (Sagan, 1995b, pp. 126–128; Waltz, 1995c, pp. 108–110). The \( \chi^2 \) test of all the data shows no statistical difference between Scenarios III and IV (when both states have few weapons, and when both states have many weapons). This tends to support Waltz’s position. When extraneous data were removed from the analysis, however, the \( \chi^2 \) test shows that the situation in which states have few weapons is statistically more likely to lead to nuclear war than the situation in which both states have many weapons.

The experimental results are also not particularly enlightening about another point of debate. The subjects’ willingness to use nuclear weapons in Scenarios II, III, and IV despite uncertainty about the outcome is an interesting result, but we should be hesitant about claiming that it counters Waltz’s argument that uncertainty leads to deterrence. Uncertainty was present to the same degree throughout the experiment, and thus the importance of uncertainty was not tested directly. Future experiments could be designed to test Waltz’s argument by controlling and/or varying the level of uncertainty.

Scenario II appears to stand out as a hazardous situation. Subjects reached the nuclear war outcome 23 percent of the time when each side in a dyad possessed a few nuclear weapons, and 10 percent of the time when each side possessed many weapons. When one side possessed few weapons and the other side possessed none—the situation most analogous to nuclear proliferation—the subjects reached the nuclear war outcome 64 percent of the time. This statistically significant result suggests that nuclear proliferation is potentially dangerous, more so when only one side in a rivalry develops weapons. While Waltz (1995b) optimistically claims that countries in this situation will not use their weapons (pp. 12, 16), this research is much more
pessimistic about the effect of nuclear proliferation.\textsuperscript{13}

The results of this research also appear to counter some of the arguments of the “Power Transition” school of international politics. For example, Kugler (1995) claims that “large disparities in power lead to peace,” (p. 12) and that nuclear proliferation is dangerous only if both sides in a dyad have nuclear weapons (pp. 24–25). To be fair, this experiment does not directly test a situation in which a weaker state challenges a stronger one (the “power transition”). Additionally, this model does not rely on the assumption that the stronger state prefers the status quo to war (Kugler and Zagare, 1990, p. 272; Zagare and Kilgour, 2000). Zagare and Kilgour use this assumption in their model of unilateral deterrence because they believe “deterrence is impossible without it” (2000, p. 138). Given the debate in the literature over the requirements of deterrence, however, it makes little sense to bias an experiment by requiring that the stronger state strictly prefers the status quo to fighting (Simon, 2002). Additionally, a model of nuclear proliferation must examine the possibility that only one side in a dyad possesses nuclear weapons.

Stepping Back: The Usefulness of this Experiment

There is no consensus in the literature on how to evaluate the validity of a simulation. Several authors discuss the external validity of simulations in broad terms, noting that it is important for simulations to relate closely to the real world (Hermann, 1967; Kelley, 1968; Macrae and Smoker, 1967, pp. 4–6; Raser, 1967, p. 62; Tanter 1972, pp. 570–572; Verba 1964). However, these authors do not provide strict criteria that could be used to test the validity of simulations. Davis (1966) notes: “Whether the results [of simulations] can be extrapolated to the real world remains arguable. Generally, it is assumed that, as a minimum, simulations sufficiently representative of the real world will produce insights and add support to (or refute) arguments of an informal, less structured nature” (p. 240). He does not, however, establish standards for knowing when a simulation is “sufficiently representative.” Even making specific statements about validity may not be a particularly illuminating enterprise. Oswalt (1993) addresses many validity issues and lists the following criteria for evaluating simulations: Affordability, assessability, data availability and data base construction, efficiency, fidelity, manpower requirements, modifiability and maintainability, portability, reliability, resulting degree of improvement, standardization, user friendliness, validity, VV&A (verification, validation, and accreditation), and verisimilitude (pp. 163–164). He discusses the relative importance of each criterion depending on the type of simulation, but does not discuss how to determine if a simulation meets a given criterion.

A trade-off exists between the sophistication of a model and its testability. Models that closely resemble the real political world are rich in detail but are very difficult to test. Models that are more abstracted from the real political world are admittedly simple, but are much easier to test.

The model presented in this article is very much abstracted from the world of international nuclear crisis bargaining. Ideally, a researcher could complement an abstract model with an experiment that is an extremely close test of the model. Unfortunately, due to the issue area being studied, there is an inherent limit to the applicability of this experiment.
Fundamentally, nuclear weapons are interesting in international politics because of the massive destruction that they can bring. Most would consider this to be the key characteristic of nuclear weapons. Yet, the most “damage” any subject suffered in the experiment was the loss of a few expected dollars. Clearly this does not approach the damage one would suffer in even the most limited nuclear attack.

This begs a difficult question: How can experimenters model death and not violate Human Subject Committee guidelines? The most severe punishment inflicted in this experiment was to have the subjects earn zero dollars for a particular round. One obvious alternative design would be to have subjects leave the experiment early if they “lost” in a war. However, that changes the game from one in which each round is independent to a more complicated game of sequential play.

Another alternative would be to inform subjects that the impact of their decisions might be “nuclear war,” rather than keeping them completely in the dark as to the purpose of the experiment. Of the few experiments on international conflict, two use this approach. The instructions to Niou and Ordeshook’s (1988) experiment make it clear that subjects were playing a “conflict game” over a “Current Threat” with an “Attacking Group” and a “Target Group” (pp. 438–439). As Pilisuk (1984) describes his own experimental setup, “[A] person’s board is called his country and the terminology of the arms race-disarmament dilemma is used” (p. 298). While it may be admirable to explicitly incorporate international elements into an experiment, a researcher runs the risk of introducing factors out of her control. To that extent, the Niou and Ordeshook and Pilisuk projects combine elements of simulations and experiments. (A more extensive future experiment could measure the impact of introducing role-playing elements, thus allowing the researcher to control for such factors. Future studies could also use people with strong leadership characteristics as subjects, with the aim of more closely representing the thinking of political decision-makers.)

In contrast, the instructions given in my game followed the example of Morgan and Wilson’s (1989) experiment on international crisis bargaining. Given the important implications of how my instructions would be received, I ran a version of the experiment in a pretest in which subjects were told they were playing a nuclear war game. However, I was unable to control their response to the term “nuclear war.” In a postexperiment interview, subjects indicated that they believed nuclear war to mean global destruction, even when the payoffs from the game they were playing indicated that one side won the war and benefited from fighting. More troubling was that the pretest took place in an undergraduate game theory course, where the students (presumably) recognized the scenarios as game trees, and the outcomes as game-theoretic payoffs. Nonetheless, labeling an outcome as “nuclear war” radically distorted the play of the game. Thus while the experimental results described here are useful as a scientific exercise, better theory and better measures are needed to reliably test statements about nuclear acquisition and international conflict.

NOTES

1. Note that the term “proliferation” implies a systemic phenomenon, namely the spread of nuclear weapons, as more countries acquire them. Previous proliferation research, however, moves down a level of
analysis, examining the effects of proliferation on a state or on a dyad. (For an exception, see Brito and Intriligator, 1996.) This study also works within the dyadic level of analysis, which makes prediction a more manageable task and follows the international relations research trend that finds state- and dyad-level explanations for international behavior (Geller and Singer, 1998).

2. Iran had a rudimentary chemical capability, which it used during the Iran–Iraq war. According to the Federation of American Scientists, Iran’s inability to fully respond to Iraq’s chemical weapons is one of the main factors behind its current chemical weapons program (www.fas.org/nuke/guide/iran/cw/index.html). See also Cordesman (1999).


4. Guetzkow and Valadez (1981, pp. 230–234) refer to a 1969 study by Raser and Crow and a 1974 study by Hermann, Hermann, and Cantor, on the effect of a related variable—the acquisition of invulnerable command-and-control for nuclear weapons. Those studies, according to Guetzkow and Valadez, found that the increased command security increased the likelihood of nuclear war.

5. Although Kugler (1998) also advocates selective proliferation, it is because “proliferation” to Kugler means that a rivalry moves from a situation in which neither state has nuclear weapons to a situation when only one has nuclear weapons.

6. For a discussion of the value of using undergraduates as subjects, see Friedman and Sunder (1994, pp. 39–43).

7. Recall that this study does not directly examine “proliferation” in the traditional use of the word—the dynamic spread of nuclear weapons. Rather, the scenarios look at “snapshots” of proliferation, by imagining four scenarios of dyadic weapons development.

8. For a formalized description of uncertainty about the outcome of nuclear war, see Goldstein (1992, pp. 524–527). In addition, this concept is not limited to nuclear proliferation. In order to enhance deterrence, Snow (1983, ch. 4) advocates modifying U.S. nuclear strategy to increase uncertainty about the outcome of war. Although Waltz points out the importance of uncertainty, elsewhere in the same chapter he argues that certainty about the outcome of war is important for deterrence (1995b, p. 6).

9. For a defense of role-playing simulation on this point, see Bradford (1995).

10. Expected Value for Player A picking right at the bottom stage = (.9)(2.7) + (.1)(1.4) = 2.57. Expected Value for Player B if Player A picked right at the bottom stage = (.9)(.3) + (.1)(1.5) = .42.

11. $\chi^2 = \sum \frac{(OBSERVED_i - EXPECTED_i)^2}{EXPECTED_i}$. See Davis and Holt (1993, 533).

12. Subjects have not behaved in strict accordance with the predictions of expected-utility and game theory in many other experiments. For a review of the literature on experiments testing why subjects behave this way, see Camerer (1995, pp. 617–676); Davis and Holt (1993, p. 508); Roth (1995a, pp. 274–288); Roth (1995b, pp. 72–86).

13. Strictly speaking, this article is not about the proliferation process; rather, it compares different scenarios of nuclear armament. One could read these results as supporting the acquisition of nuclear capability by a non-nuclear state in an asymmetric dyad. This is consistent with Bueno de Mesquita and Riker’s second model (1982, p. 300).

REFERENCES


**CONTRIBUTOR**

**Michael W. Simon** is Chief of the International Relations Branch at SPARTA, Inc. His research focuses on terrorism, proliferation, nuclear forces, missile defense, and game theory.