International Studies Review (2018) **20**, 422–437

ANALYTICAL ESSAY

Explaining the Long Peace: Why von Neumann (and Schelling) Got It Wrong

Frank C. Zagare
Department of Political Science, University at Buffalo, State University of New York, Buffalo, NY 14260, USA

Alexander J. Field (2014, 54) argues that game theory "offers little guidance, normatively or predictively, in thinking about behavior or strategy in a world of potential conflict." He makes this claim by attributing to John von Neumann a view of the superpower relationship during the Cold War period that has no basis in fact and inferring policy prescriptions to that view that are simply not there. Field also suggests that Thomas Schelling's explanation of the "event that didn't occur" leads to the conclusion that "deterrence works because we are human, not because we are entirely rational" (Field 2014, 86). In this essay I show that there is at least one logically consistent game-theoretic explanation of the absence of a nuclear war during the long-peace of the 1950s and early 1960s. I also demonstrate that Field's assumptions lead to exactly the opposite conclusions; that is, that mutual deterrence can in fact be reconciled with rationality and that game theory is a powerful tool for understanding interstate conflict.

Keywords: Deterrence, Game Theory, Schelling, von Neumann, Rationality, Cold War

Introduction

Game theory is best thought of as a logical system. Technically speaking, it is a branch of mathematics. As a distinct field of study, game theory entered the academic world in 1944 when John von Neumann and Oskar Morgenstern published their magisterial *Theory of Games and Economic Behavior* with Princeton University Press. Its broad acceptance as a legitimate methodological tool and its application to tactical military affairs was almost immediate (see, inter alia, McDonald 1950; McDonald and Tukey 1949; Williams 1954). There were few objections. But when, shortly thereafter, game-theoretic models found wide currency among *strategic* analysts and international relations theorists, there was a backlash, both from leading game theorists—including Morgenstern (1961) himself and Anatol Rapoport (1964), one of the most prominent game theorists of his time—and from more traditional scholars who objected to treating questions of war and peace as a mere "game" (Zuckerman 1956).

After the publication of Thomas Schelling's most influential book, *The Strategy of Conflict*, in 1960, applications of game theory to national security affairs all but

Zagare, Frank C. (2017) Explaining the Long Peace: Why von Neumann (and Schelling) Got It Wrong. *International Studies Review*, doi: 10.1093/isr/vix057

disappeared in the literature of international relations. There were, of course, a few exceptions. But for the most part strategic analysts concluded that Schelling had said all that could be said by a game theorist about deterrence in general and coercive bargaining in particular (Martin 1999).

Toward the end of the 1970s and the beginning of the 1980s, however, applications of game theory began to reappear in some of the more specialized political science journals. At first, it was a trickle. But by the end of the century, game models came to be accepted as part of the theoretical mainstream.[1] Predictably, there was another backlash (e.g., Johnson 1997). One critic even accused formal modelers in general, but game theorists in particular, of hegemonic aspirations in the field of security studies and warned that the increasing dominance of game-theoretic studies threatened to calcify the field by privileging some questions over others (Walt 1999). Not surprisingly, game theorists disagreed (Bueno de Mesquita and Morrow 1999; Martin 1999; Niou and Ordeshook 1999; Powell 1999; Zagare 1999).

More recently, the utility of game-theoretic models for understanding interstate behavior has been challenged by some behavioral economists (a.k.a. cognitive psychologists), who contend that real world decision makers, who suffer from a number of cognitive and motivated biases, are not "rational" in a game-theoretic sense (Carlson and Dacey 2006, 2014; Levy 1997; McDermott 2004). The argument has significant implications. If one accepts it, as does Alexander J. Field (2014, 54), it follows that game theory "offers little guidance, normatively or predictively, in thinking about behavior in a world of potential conflict."

My purpose in this essay is to dispute Field's argument—and by extension, the conclusion that some other behavioral economists reach—by demonstrating that, properly understood, game theory is a powerful tool for understanding conflict behavior, interstate or otherwise. My purpose, however, is not to contest the significant insights behavioral economists have uncovered about human decision-making. It is my contention that their insights and empirical observations are not necessarily inconsistent with the gestalt of game theory. As already mentioned, game theory should be understood as nothing more and nothing less than a potentially useful *methodology* for understanding human choice in an interactive decision-making environment (Morton 1999). It is, therefore, one thing to find fault with an *application* of game theory or an *argument* that a game theorist might make. It is quite a different thing to conclude that the methodology itself is at fault. For example, if a bridge fails or if a building collapses because of improper design, it does not follow that engineering as a field has nothing to contribute to the design of, say, physical objects. So it is with game theory. If a game theoretic study is either logically inconsistent or empirically inaccurate, clearly the study should be rejected; at the same time, it should also be obvious that the same is true of any theoretical argument regardless of its microfoundation.

To develop this point, I begin by exploring Field's (2014) attempt to explain why the United States and the Soviet Union were able to avoid an all-out thermonuclear exchange during the Cold War era. His main conclusion is that deterrence worked "not because we are entirely rational" but "because we are human" (Field 2014, 86). In so concluding, Field explicitly accepts a game-theoretic interpretation of the superpower relationship that he attributes to John von Neumann and argues that inconsistencies in the work of Thomas Schelling reinforce his argument that a rationalist explanation of what Gaddis (1986) calls the "long peace" of the Cold War years is not possible.

Field's argument, however, cannot be sustained. Briefly, I argue that his interpretation of von Neumann's conceptualization of the game played between two nuclear adversaries, coupled with his own description of the dynamic nature of that game, leads to the exact opposite conclusion. Moreover, I show that Schelling's explana-

[1] Zagare and Slantchev (2012) trace the development of game-theoretic applications in international relations.

tion of the absence of a superpower conflict during the Cold War period falls apart both logically and empirically. In consequence, so does Field's.

It should be noted at the outset that Field's (2014, 55–56, n8) argument begins with an overly rigid definition of what constitutes rationality. In addition to the standard axioms that are associated with a von Neumann-Morgenstern (1944) cardinal utility function, Field asserts that rationality requires agents who are not only self-interested but also self-regarding.[2] As Field recognizes, neither altruists nor suicide bombers are self-regarding (by his definition) and, therefore, cannot be considered rational if this restriction on preferences is accepted. Thus, his nonstandard definition[3] unnecessarily places a great deal of human behavior outside the realm of scientific exploration within a rational choice framework. But even if this were not the case, to claim, as Field (2014, 57) does, that "the only defensible policy" for a self-regarding agent in a contentious bilateral nuclear relationship "was an immediate attack" that would bring about the destruction of tens and hundreds of millions of innocents is to offer a caricature of the very notion of rational choice and self-regarding behavior. As Field reports, during the Eisenhower years there were indeed some rational, perhaps even self-regarding, individuals who pushed for a preemptive attack on the Soviet Union. Fortunately, there were also some rational self-regarding individuals, including Eisenhower himself, who saw things differently.

The point here is not that Field's definition of rationality is wrong; rather, that it is overly demanding and, therefore, unnecessarily restrictive.[4] Nonetheless, Field's argument does not fail because of it but in spite of it as I next demonstrate. In other words, his conceptualization of rationality is not the reason his argument is less than persuasive.[5]

John von Neumann and the Prisoners' Dilemma

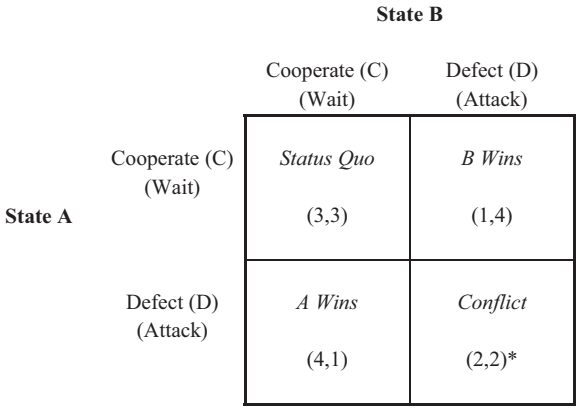
Field builds his case against game theory, rationality, and von Neumann's obviously incorrect fear that a nuclear exchange between the superpowers was all but certain by attributing to von Neumann the view that nuclear confrontations are essentially prisoners' dilemma games. Field (2014, 54, n4) claims that his attribution is "indisputable." But it is far from it. In fact, the primary source that Field uses to

[2] For Field, "self-regarding" behavior requires that individuals "prefer more material goods to less, and life over death," *whatever the consequences or collateral implications*. Field claims that game theory can be tested only when preferences are so restricted. This would be the case if game theory is taken to be a descriptive—as opposed to a normative—framework. But game theory can also be considered a tool for theory construction. In other words, once preferences are postulated and strategic choices stipulated, a game-theoretic model can lead to testable propositions. Downs (1957), for example, assumes that political candidates and/or parties are rational vote maximizers. This assumption leads to specific behavioral expectations that are subject to empirical validation. Behavior inconsistent with these expectations undermines his theory of electoral competition.

[3] A von Neumann-Morgenstern utility function is a measure of an actor's *subjective* preference over outcomes given uncertainty. Therefore, individuals with different preferences and risk propensities may have distinct utility functions. In other words, utility is defined by each individual. Field (2014, 74) misstates the facts, then, when he claims that game theorists commonly assume "that players are logical, rational, *and self-regarding*" (emphasis added) as he defines it. Self-regarding behavior, like beauty, is in the eye of the beholder (Kreps 1988). So even if von Neumann considered a preemptive nuclear attack on the Soviet Union to be self-regarding, another (also) rational agent might think otherwise.

[4] Definitions, by their very nature, are arbitrary. In the deterrence literature, two definitions of rationality figure prominently. The concept of *procedural rationality* underlies the work of those who approach strategic behavior from the vantage point of individual psychology (Simon 1976). Most rational choice deterrence theorists who study deterrence define rationality instrumentally. For a further discussion, see Zagare (1990) and Quackenbush (2004). Schelling (1960, 1966) is inconsistent; he uses both. His "rationality of irrationality" stratagem assumes that instrumentally rational agents feign procedural irrationality to gain a bargaining advantage.

[5] It should also be noted that once preferences are stipulated, the distinction between rational and self-regarding behavior disappears. Behavior that is "rational" and behavior that is "self-regarding" are one and the same; that is, behavior that is, and must be, consistent with a player's utility function. As will be seen, Field seems to understand this when analyzing's Schelling's explanation of the absence of a nuclear war throughout the Cold War period but not when he discusses what he believes to be von Neumann's understanding of the game played by the United States and the Soviet Union.



Key: (x,y) = payoff to State A, payoff to State B
4 = best; 3 = next-best; 2 = next-worst; 1 = worst
* = Nash equilibrium

Figure 1. Prisoners' Dilemma

make this case, William Poundstone (1982, 144), argues otherwise: "It's unlikely that von Neumann—or anyone else, circa 1950—explicitly thought of the U.S.-Soviet conflict as a prisoner's dilemma. If von Neumann did picture U.S.-Soviet relations as a game, it is more plausible that he saw it as a zero-sum game."[6]

For the sake of argument, however, let us accept Field's less-than-convincing contention that von Neumann saw the superpower relationship as a prisoners' dilemma game. A standard depiction of this well-known game is given by Figure 1. Prisoners' dilemma is a two-person noncooperative game. In this representation, the players are nuclear adversaries, here called State A and State B. Each state has two strategies, to cooperate (C) or to defect (D). There are four possible outcomes: if both players choose to cooperate and do not attack one another, the *Status Quo* results; if both defect and attack, *Conflict* (read nuclear war) takes place. But if one defects and attacks while the other cooperates (by waiting to attack), the state that attacks *Wins* and the state that waits *Loses*.

Prisoners' dilemma is a 2 × 2 normal-form game. There are seventy-eight such games (Rapoport and Guyer 1966), but the preference assumptions that define this game are unique: both players are assumed to most-prefer to *Win*, second-most prefer the *Status Quo*, third-most prefer *Conflict*, and least prefer to *Lose*. In Figure 1 these preference assumptions are indicated by an ordered pair in each cell of the matrix, where the first entry represents State A's (i.e., row's) ranking and where the second entry represents State B's (i.e., column's) ranking of the associated outcome. The outcomes are ranked from best (i.e., four) to worst (i.e., one). For example, the outcome called *A Wins* is best for State A and worst for State B.

As is well known, both players in a prisoners' dilemma game have a strictly dominant strategy; that is, a strategy that is best regardless of the strategy selected by the other player. For instance, State A's D (attack) strategy is A's best response should State B attack. But is also State A's best response should State B wait to attack. Similarly, State B's attack strategy is a best response to either of State A's two strategy choices.

Strictly dominant strategies, in other words, are unconditionally best. Thus "rational" players, self-regarding or otherwise, will always select them. And when they do,

[6] If Poundstone is correct, Field must be wrong. Prisoners' dilemma is not a zero-sum game.

they lead to a unique, non-Pareto optimal Nash equilibrium, *Conflict*. Nash equilibrium is the standard measure of rational play in normal-form games. Game theorists are in almost unanimous agreement that in a one-shot game, which a thermonuclear war would surely be,[7] rational players should select their dominant strategies and that *Conflict* is the outcome that is implied under rational play. Of course, there is a dilemma. In this game, two rational players are individually and collectively worse off than two irrational players who choose to cooperate. Unfortunately, in a one-shot Prisoners' Dilemma game, mutual cooperation cannot be supported under rational play.[8]

Although it is difficult to reconstruct his logic, Field's (2014, 54) argument that game theory "offers little guidance, normatively or predictively, in thinking about behavior or strategy in a world of potential conflict" rests on several indisputable facts: 1) von Neumann believed that an all-out nuclear war between the United States and the Soviet Union was inevitable and that, therefore, it was to the advantage of the United States to attack first; 2) rational players in a prisoners' dilemma game should select their dominant strategy, and, if and when they do, conflict is implied; 3) a thermonuclear exchange between the United States and the Soviet Union did not occur; 4) there is a disconnect between von Neumann's belief, his policy recommendation, the strategic imperatives of the prisoners' dilemma game, and "the event that didn't occur."

But his argument also rests on some facts that are more than disputable: 1) von Neumann believed that the game played by the superpowers during the earliest years of the Cold War was indeed a prisoners' dilemma game; 2) "von Neumann was right to characterize … [the superpower relationship as] … a Prisoners' Dilemma" (Field 2014, 55); and 3) that von Neumann's preferred policy (i.e., a *preemptive* strike) is supported by his belief.

As already noted, Field's assertion that von Neumann believed that a prisoners' dilemma game captured the underlying dynamic of the post war relationship of the United States and the Soviet Union is speculative at best and factually incorrect at worst. On the other hand, von Neumann's beliefs are simply beside the point.[9] So even if Field's claim is accepted, the pertinent questions are 1) whether or not the Cold War competition between the superpowers was, in fact, a prisoners' dilemma game; and 2) assuming for the sake of argument that it was, is von Neumann's preference for a *preemptive* attack consistent with the conventional wisdom of game theory?

It is clear that the answer to the second question is "no." As noted, prisoners' dilemma is a 2×2 normal-form game in which each player has two strategies. In a normal-form representation, the assumption is that the players make their strategy choice before the game begins. There are two logically equivalent interpretations of a 2×2 game. One is that the players choose their strategies simultaneously; the second is that they make their choices sequentially but without knowledge of each other's choice. In a 2×2 normal-form game, therefore, there can be no first-mover advantage since there is, technically speaking, no first mover. And if there is no first mover, there is also no second mover and, hence, no possibility of retaliation. Thus,

[7] To say that a game whose unique equilibrium is associated with an all-out nuclear war is a one-shot game is not the same thing as saying that the superpower relationship during the Cold War can be captured by the strategic dynamic of a prisoners' dilemma game. A one-shot game is any dynamic (i.e., extensive-form) or static (i.e., strategic-form) game played once. The final game of any finite repeated game can be considered a one-shot game (Luce and Raiffa 1957, 100). It is in this sense that an ongoing deterrence game that culminates in thermonuclear war can be thought of as a one-shot game.

[8] In a repeated game, there are conditions under which mutual cooperation is part of a Nash equilibrium. For a discussion, see Morrow (1994, 260–62).

[9] They are beside the point because even if Field is correct about von Neumann's beliefs, Field's conclusions about the usefulness of game theory as either a normative theory or as a tool for theory construction do not hold, as demonstrated below.

		State B:			
		C/C *C Regardless*	*D/D* *D Regardless*	*C/D* *Tit-for-Tat*	*D/C* *Tat-for-Tit*
State A:	*C*	(3,3)	(1,4)	(3,3)	(1,4)
	D	(4,1)	(2,2)*	(2,2)	(4,1)

Key: (x,y) = payoff to A, payoff to B
4 = best; 3 = next-best; 2 = next-worst; 1 = worst
* = Nash Equilibrium

Figure 2. Sequential Prisoners' Dilemma

either von Neumann was not a very good game theorist or he had a different game in mind.[10]

It is important to point out that there also is no first-mover advantage in a prisoners' dilemma game even if it is played sequentially. A normal-form representation of such a game is depicted in Figure 2. In this representation, the assumption is that State A chooses its strategy first and then, after observing this choice, State B makes its choice. In the sequential version of this game, State A's strategy set is the same, but now State B has four strategies:

1. "C Regardless"—choose C if State A chooses C/choose C if State A chooses D (C/C),
2. "D Regardless"—choose D if State A chooses C/choose D if State A chooses D (D/D),
3. "Tit-for-Tat"—choose C if State A chooses C/choose D if State A chooses D (C/D), and
4. "Tat-for-Tit"—choose D if State A chooses C/choose C if State A chooses D (D/C).

In the sequential version of prisoners' dilemma depicted in Figure 2, State B's D Regardless strategy weakly dominates all of its other strategies. And since State A's best response to B's weakly dominant strategy is to choose D initially, the strategy pair (D; D/D) gives rise to the unique, Pareto-inferior Nash equilibrium (2,2), or *Conflict*. All of which is to say that there is no discernable strategic difference between a prisoners' dilemma game played simultaneously or played sequentially. There is no first mover advantage in either version of the game.

It should be clear, then, that von Neumann's policy preference for a preemptive strike against the Soviet Union cannot be logically derived from a one-shot prisoners' dilemma game, however it is played out. In either the simultaneous or sequential case, the fact that the United States did not preempt the Soviet Union at the dawn of the nuclear era does not support Field's (2014, 54) observation that "game

[10] He would not be a good game theorist because his argument for a first strike cannot be deduced from the strategic implications of a prisoners' dilemma game. He might have a different game in mind because there are in fact some games (e.g., chicken—see below) in which there is a first-mover advantage.

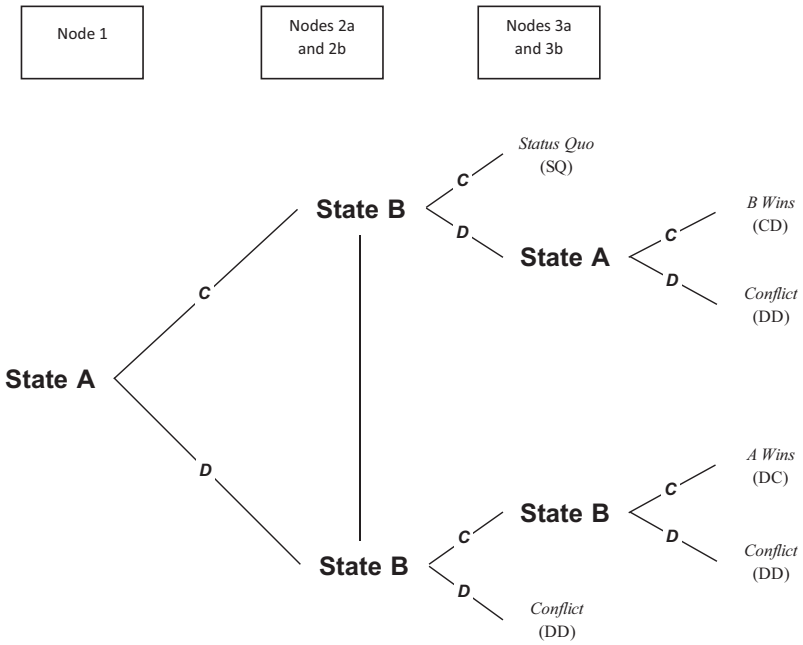


Figure 3. Generalized Mutual Deterrence Game

theory leads to behavioral predictions which are simply not borne out … in the real world."[11]

Rational Deterrence

All of which suggests that the standard version of prisoners' dilemma is not a very good model of the superpower relationship at the time that von Neumann recommended a preemptive attack. Assuming, then, that von Neumann was in fact a good game theorist, it follows that he must have had another game in mind. Which one, however, is unclear. The extensive-form game of Figure 3 is, at a minimum, a plausible possibility. As with the standard version of prisoners' dilemma, this game, which Zagare and Kilgour (2000) call the "generalized mutual deterrence game," is symmetric. As well, both players are presented with the same *initial* choice that the players in a prisoners' dilemma game have, to either Cooperate (C) or to Defect (D). Both players are also assumed to make their initial choices simultaneously or, what is an equivalent assumption, without knowledge of the initial choice of the other player. The generalized mutual deterrence game, however, allows for the possibility of a retaliatory strike if initially only one player chooses to defect by attacking. The possibility of retaliation is the *only* difference between the rules that govern play

[11] Much the same could be said about the experimental literature that suggests that players in prisoners' dilemma games do not always defect (Schecter and Gintis 2016, 13). If the standard version of prisoners' dilemma is an inappropriate model of the Cold War interaction of the superpowers, then the laboratory experiments that explore behavioral tendencies in an artificial environment are simply beside the point, and the conclusion that Field (2014, 86) draws from that literature, that "formal game theory has not been useful for understanding how people behave or how they necessarily should behave," is not germane to his analysis of the predictive or explanatory power of game theory in the field of security studies.

State A \ State B	C/C	C/D	D/C	D/D
C/C	(3,3)	(3,3)	(1,4)	(1,4)
C/D	(3,3)	(3,3)**	(2,2)	(2,2)
D/C	(4,1)	(2,2)	(2,2)*	(2,2)*
D/D	(4,1)	(2,2)	(2,2)*	(2,2)**

Key: * = Nash Equilibrium
** = Subgame Perfect Nash Equilibrium

Figure 4. Ordinal Strategic-Form Representation of the Generalized Mutual Deterrence Game Given Prisoners' Dilemma-like Preferences

in the generalized mutual deterrence game and those that are associated with the standard version of prisoners' dilemma. The claim here is that the dynamic structure of the extensive-form game of Figure 3 better captures the strategic situation envisioned by both von Neumann and by Field.

There are four distinct outcomes in the generalized mutual deterrence game. They are identical to the four outcomes in prisoners' dilemma. If both players cooperate, the *Status Quo* obtains. If both defect, either initially or subsequently, *Conflict* results. But if one defects initially and the other, after initially choosing to cooperate, does not retaliate, the defecting player wins and the cooperating player loses.

To retain as much of the structure of the game as Field claims that von Neumann had in mind when he pushed for a preemptive strike of the Soviet Union, I now analyze the generalized mutual deterrence game using the same ordinal preference assumptions over the four distinct outcomes that define prisoners' dilemma. Specifically, the assumption here is that both players most-prefer to win, next most-prefer mutual cooperation, third-most-prefer *Conflict* and least-prefer to lose. The strategic form of this version of the generalized mutual deterrence game is given in Figure 4.

Notice that there are five Nash equilibria in the normal-form game of Figure 4. But only two of them are subgame perfect (Selten 1975). Subgame perfect Nash equilibrium is the accepted standard of rational play in an extensive-form game such as the one depicted in Figure 3. Unlike Nash equilibria—which may be supported by threats that are irrational to carry out and are, therefore, not credible—subgame perfect equilibria require that the players plan to choose rationally at every node of the game tree, whether they expect to reach a particular node or not. Subgame perfect equilibrium, in other words, is a refinement of Nash equilibrium and, as such, constitutes a more nuanced understanding of the requirements of rational play.

Interestingly, one of the subgame perfect equilibria results when both players intend to choose D (i.e., attack) whenever it is their turn to make a move. A D Regardless strategy in the generalized mutual deterrence game is, in essence, the same as a D strategy in the standard 2 × 2 version of prisoners' dilemma and, not

surprisingly, results in the same outcome, *Conflict*, which is the next-worse outcome for both players, that is, (2,2).

Significantly, however, the second subgame perfect equilibrium occurs when both players cooperate initially but intend to defect (retaliate) subsequently should the other player defect initially. Thus, a tit-for-tat strategy of conditional cooperation that justifies a nonpreemptive strategy is also fully consistent with the requirements of rational choice. As well, unlike the *Conflict* subgame perfect equilibrium, the subgame perfect equilibrium that maintains the *Status Quo* is Pareto-optimal. Thus, both self-interested and self-regarding players have a very good reason to select the strategies that bring it about.

All of which is to say that even a slight modification of the game that Field contends drove von Neumann's thinking produces another rational strategic possibility that one could plausibly argue characterized the Cold War period. It is noteworthy that the modification that produced the generalized mutual deterrence game is clearly closer to Field's (2014, 76) and von Neumann's description of a contentious nuclear relationship. Recall that the condition that Field (2014, 57) used to justify von Neumann's view of the game included the possibility of retaliation. The generalized mutual deterrence game analyzed in Figures 4 and 5 incorporates not only von Neumann's presumed understanding of the game's payoff structure but the possibility of a retaliatory choice as well. In a 2×2 normal-form representation of the prisoners' dilemma, where the assumption is that players select their strategy choices simultaneously (or in ignorance of each other's choice) before the play of the game, there is no possibility of a retaliatory strike.

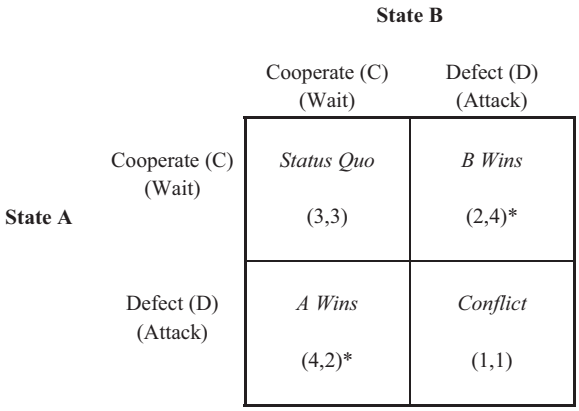
Still, one might object. As Field (2014, 55) notes, his acceptance of von Neumann's characterization of contentious nuclear relationships required that there be "some possibility of destroying an adversary's offensive capability and/or its will to retaliate." There is no such possibility in the complete information version of the generalized mutual deterrence game. Given prisoners' dilemma–like preferences, rational players *always* retaliate.

One way to model the *possibility* that one or both of the players might not be able to retaliate is to consider the equilibrium structure of the generalized mutual deterrence game *with incomplete information*.[12] When it is, the conclusions that can be drawn from an examination of the game under complete information are not disturbed. As Zagare and Kilgour (2000, 111–21) show, there are three distinct perfect Bayesian equilibria in the incomplete information version of the generalized mutual deterrence game that are rationally consistent with maintenance of the status quo and an initially cooperative strategic stance.[13] One of them, which they call the sure-thing deterrence equilibrium, *always* leads to the status quo regardless of the players' actual preference between retaliating or not, that is, the players' types.

All of which is to say that it is difficult to agree with Field's (2014, 54) argument that game theory's "canonical behavioral assumptions predicted devastating conflict between nuclear adversaries," even though von Neumann did. When the preference assumptions associated with the standard version of the prisoners' dilemma game are maintained and used to analyze an extensive-form game that more fully reflects von Neumann's understanding of the dynamics of deterrence, a devastating conflict remains but one of two possible rational strategic possibilities, and one that self-regarding players would reject at that. Thus, the predictive failure that Field attributes to game theory does not exist.

[12] Under incomplete information the players do not know each other's preferences. Thus, the assumption of incomplete information is logically equivalent to the assumption that one player believes that it is a distinct *possibility* that the other player will choose not to retaliate or, equivalently, will be incapable of retaliating effectively.

[13] A perfect Bayesian equilibrium is the standard measure of rational play in an extensive-form game with incomplete information. A perfect Bayesian equilibrium specifies an action choice for every type of every player at every decision node or information set belonging to the player; it must also indicate how each player updates its beliefs about the other player's type in the light of new information obtained as the game is played out.



Key: (x,y) = payoff to State A, payoff to State B
4 = best; 3 = next-best; 2 = next-worst; 1 = worst
* = Nash equilibrium

Figure 5. Chicken

Thomas Schelling and Chicken

According to Field, von Neumann's policy prescription was opposed most forcefully by Thomas Schelling. But if Schelling wanted to develop a strategic rationale for a policy of deterrence, he chose a strange game form to do so. His starting point—and the starting point of the vast majority of those strategic thinkers who used game theory to study deterrence—was the game of chicken (see Figure 5) (Schelling 1966, 116–25).

In some respects, chicken resembles game theory's other canonical game and the purported object of von Neumann's attention, prisoners' dilemma. Chicken, like prisoners' dilemma, is a 2 × 2 normal-form game in which each of the two players has two strategies, either to Cooperate (C) or to Defect (D) from cooperation. As is the case in prisoners' dilemma, each player is assumed to most prefer to win and second most-prefer to cooperate when the other player also cooperates. The two games diverge, however, with respect to the relative ranking of the two remaining outcomes. In chicken, *Conflict* is a mutually worst outcome, and losing is second-worst. In prisoners' dilemma, these preferences are reversed.

The difference is not trivial. In prisoners' dilemma, *Conflict* is the unique Nash equilibrium. By contrast there are two (subgame perfect) Nash equilibria in chicken. As indicated by the asterisks in Figure 5, one is associated with a victory for the row player (State A) and the other with a victory for column (State B).[14] Significantly, the *Status Quo* (or mutual cooperation) is not a Nash equilibrium in chicken, which, on its face, presents a challenge to those who want to explain why two rational agents might choose to cooperate. On the other hand, rational *Conflict* is also ruled out once the assumption is that it is each actor's least-preferred outcome. A mutually worst outcome can never be part of a (pure) strategy Nash equilibrium in a strictly ordinal normal-form game.

[14] There is also a mixed strategy equilibrium that, for the sake of brevity, will be ignored. Under the mixed strategy equilibrium, deterrence succeeds sometimes but not necessarily often. The mixed strategy equilibrium in chicken fails as a normative device. As O'Neill (1992: 471–72) shows, it prescribes behavior that is "just the opposite of what one would expect." Under the mixed strategy equilibrium in chicken, the worse the *Conflict* outcome is for one player, the *more* likely it is that the other player will concede. Like the unique pure strategy Nash equilibrium in prisoners' dilemma, the mixed strategy Nash equilibrium in chicken is not Pareto-optimal.

		State B:			
		C/C *C Regardless*	*D/D* *D Regardless*	*C/D* *Tit-for-Tat*	*D/C* *Tat-for-Tit*
State A:	*C*	(3,3)	(2,4)*	(3,3)	(2,4)
	D	(4,2)*	(1,1)	(1,1)	(4,2)**

Key: (x,y) = payoff to A, payoff to B
4 = best; 3 = next-best; 2 = next-worst; 1 = worst
* = Nash Equilibria
** = Subgame Perfect Nash Equilibrium

Figure 6. Sequential Chicken

Making Schelling's task of justifying a nonpreemptive policy even more difficult is the fact that there is indeed a first move advantage in chicken, as Figure 6 shows. There are three Nash equilibria in sequential chicken. As indicated by the asterisks, two correspond to equilibria in the original (simultaneous choice) game while the third—(D, D/C)—is strictly a product of a sequential structure. But this additional equilibrium has special properties that distinguish it from the other two and, therefore, give it a singular status: it alone is subgame perfect; as well, it is the product of A's best-response to B's weakly dominant strategy (i.e., D/C). It is more than significant that under this equilibrium State A, the player assumed to move first, wins.

Schelling was clearly aware of the fact that in sequential chicken, the player who moves first always wins. In one of the most cited chapters of his book *Arms and Influence* entitled "The Art of Commitment," Schelling (1966) prescribed a variety of stratagems for statesmen who wish to seize the initiative and prevail in a nuclear crisis by forcing a rational opponent to "chicken out." He was not the only defense intellectual to do so. Herman Kahn (1962), Jervis (1972), Glenn Snyder (1971), and several others also trafficked in this dangerous drug.

Schelling's musings with respect to what George (1993) calls "forceful persuasion" and what Field (2014, 56, n10) refers to as "aggressive" deterrence have little basis in fact. For instance, the multiple bargaining tactics that Schelling proposed in *Arms and Influence* have been shown in more than one large-*n* study to be empirically dubious (e.g., Huth [1999]; Danilovic [2002]). And, as Jervis (1988, 80) has pointed out, in the specific case of the Cold War interaction of the superpowers, the reality is that "the United States and the USSR have not behaved like reckless teenagers" in a game of chicken. Given the discrepancy between theory and fact, it is no small wonder then that Jervis (1979, 292) asserted that many of Schelling's policy prescriptions were "contrary to common sense," that Rapoport (1992, 482) found them to be just plain "bizarre," and that Morgenstern (1961, 105) concluded that they would be "dangerous should they have an influence on policy."

Field (2014, 61, n21), who was well-aware of the empirical shortcomings of Schelling's work on coercive bargaining, thought it significant that Schelling "had a strong influence on academic thinking about strategic policy" but "limited influence" on actual policy makers in either Washington or Moscow. For Field (2014, 54) this was contributing evidence that formal game theory itself "offers little guidance, normatively or predictively, in thinking about behavior or strategy in a world of potential conflict."

But it was actually Schelling's work on what he termed "traditional" deterrence that Field (2014, 56, n10) believed conclusively demonstrated game theory's inability to adequately explain the long peace. If chicken was, in fact, the proper game form to represent the strategic relationship of two nuclear adversaries, as the majority of strategic analysts thought, and if preemption was rational in chicken, as Schelling and others had properly inferred, how might mutual deterrence evolve (or be explained)?

To answer this question Schelling abandoned both logical consistency and the rationality postulate. His response was that nuclear deterrence would only work if an aggressor was convinced that its opponent would retaliate—*irrationally*. As he put it so succinctly: "another paradox of deterrence is that it does not always help to be, or to be believed to be, fully rational, cool headed, and in control of one's country" (Schelling 1966, 37). In other words, it would be rational to be thought of as irrational. Schelling, of course, was not the first strategic thinker to play fast and loose with the rationality postulate. Bernard Brodie (1959, 293), considered by many to be the seminal deterrence theorist, put it this way: "For the sake of deterrence before hostilities, the enemy must expect us to be vindictive and irrational if he attacks us."

If for Schelling each superpower was able to deter the other only by threatening to retaliate, irrationally, how can one explain the fact that neither superpower followed von Neumann's advice and preempted the other? The only defensible answer for Schelling was that both were *rationally* deterred. In other words, Schelling's explanation assumes, simultaneously, that decision-makers in the United States and the Soviet Union were rational when they were being deterred and irrational when they were deterring one another. Schelling's inconsistent application of the rationality postulate is more than problematic. If players can be either rational or irrational, *any* course of action can be justified or explained away. Walt (1999) notwithstanding, logical consistency is clearly the sine qua non of sound theory. Since any proposition and its opposite can be derived from a logically inconsistent theoretical framework, empirical validation is foreclosed. And, as Field (2014, 56, n8, 61) correctly argues, empirical validation is a cornerstone of scientific inquiry.

Field (2014, 71, n43) saw the logical problem with Schelling's explanation and dismissed it because, as he saw it, Schelling assumed players who were not self-regarding. Recall that for Field (2014, 55, n6, 7), a self-regarding actor would rationally attack its opponent since a self-regarding opponent would *not* retaliate ex post. Schelling's explanation assumed the opposite. What is both remarkable and significant is that Field is absolutely correct—if the game is chicken and not prisoners' dilemma. In *chicken*, a self-regarding player would not retaliate (because retaliation would lead to its worst outcome and not retaliating would lead to its second-worst outcome), which is why a self-regarding player would also have an incentive to strike first (see Figure 6).

Notice that when Field dismissed Schelling's logic, he simply reversed it. Schelling argued that each superpower was rationally deterred by the threat of irrational retaliation. Field argued that each superpower was irrationally deterred by a threat of rational nonretaliation. His argument, in other words, was that it was irrational for either superpower not to attack the other since it would have been rational for the other not to retaliate. In the end Schelling and Field (2014, 79) both conclude that "nonaggressive deterrence (peaceful coexistence) requires a combination of irrational and rational behavior on the part of both parties." All of which helps to explain Field's (2014, 55, n7) claim that his "paper argues (and Schelling suggests the same), that deterrence worked and von Neumann's predictions failed because humans are not entirely self-regarding."

It is no inconsequential fact, however, that the actual incentive structure of chicken is entirely consistent with von Neumann's policy prescription. So if von Neumann was a good game theorist and if he had any game in mind when he

campaigned for a preemptive strike of the Soviet Union, it was probably chicken and not prisoners' dilemma. In prisoners' dilemma, as we have seen, a self-regarding player will in fact retaliate because retaliation will result in its next-worst outcome, but nonretaliation will bring about its worst outcome (see Figure 2). Moreover, when both players are afforded an equal opportunity to retaliate, mutual deterrence is entirely consistent with the rationality assumption (see the discussion of the generalized mutual deterrence game in Section 3)[15], and, not insignificantly, self-regarding behavior as is clearly implied by the payoff assumptions. To argue otherwise, as Field does, is to ignore the incentive structure implicit in any game with preferences that mirror those of a standard version of a prisoners' dilemma game.

All of which is to say that Field's assumption that the strategic relationship of the superpowers during the Cold War period is captured by the payoff structure of a prisoners' dilemma game—and that neither player possessed a first-strike capability that eliminated the possibility of a retaliatory strike—leads to a conclusion that is exactly the opposite of his. More specifically, if the superpowers were in fact involved in a mutual deterrence relationship,[16] then deterrence worked not only because we are human but also because we were rational (Zagare 2004). It also follows, contrary to Field and some other behavioral economists, that game theory indeed has much to offer, both normatively and descriptively, about behavior and strategy in a world of intense interstate conflict, even though von Neumann and Schelling (and Field) got it wrong.

Coda

In a far-ranging article, Alexander J. Field makes the argument that during the Cold War period, peace was preserved not because the superpowers were rational but because they were human. He claims, without convincing evidence, that John von Neumann, one of the cofounders of game theory, saw the superpower relationship during that period as a prisoners' dilemma game and contends that von Neumann's policy preference for a preemptive nuclear attack of the Soviet Union by the United States, while game-theoretically sound, was empirically inaccurate. By contrast, he argued that Thomas Schelling's explanation of the long-peace was game theoretically faulty but empirically correct. Field (2014, 54) concludes that "game theory leads to behavioral predictions which are simply not born out in the laboratory or … in the real world."

In this article, I argue that Field's inference about the rational basis for a *preemptive* attack in a prisoners' dilemma game is not logically supported. I also show that a slight modification of that games' strategic structure that more closely mirrors the dynamic decision-making context that both von Neumann and Field have in mind leads to exactly the opposite conclusion: namely, that self-regarding players with prisoners' dilemma-like preferences might rationally be deterred. Finally, I show that Field's analysis of Schelling's explanation of the long peace, which is game-theoretically sound, is at odds with his analysis of the game he attributes to von Neumann and that he accepts as the proper representation of a contentious nuclear relationship.

[15] To say that there are conditions under which mutual deterrence is consistent with rational choice is not the same thing as saying that mutual deterrence is either robust or all-but-certain. For the argument that mutual deterrence is both rational and fragile, see Zagare (2011, 39–57).

[16] It is also possible that one side or the other was not incentivized to upset the status quo. In this case the relationship would be one of unilateral deterrence. [For the argument that during the Cold War period the United States was a satisfied status quo power, see Organski and Kugler (1980)]. Under specific conditions, unilateral deterrence relationships are also consistent with rational choice (Zagare and Kilgour, 2000). Of course, if neither side preferred to upset the existing order, deterrence was not germane. It remains an important empirical puzzle whether the long peace was an instance of unilateral or mutual deterrence or if deterrence was even relevant.

The larger point, however, is that Field's outright rejection of game theory as an explanatory or predictive tool cannot be sustained. Game theory constitutes a powerful methodological tool for theory construction. Even if one accepts Field's argument that von Neumann's policy preference for a preemptive attack on the Soviet Union was game-theoretically correct, and that his prediction based on it was empirically inaccurate, it does not follow that formal game theory is of limited utility "for thinking about behavior or strategy in a world of potential conflict" (Field 2014, 54).

Game-theoretic models should be thought of as empty vessels that can be filled in or even shaped by conflict theorists. As Morrow (1994, 70) points out, the specification of a model is the single most important step in the development of theory. What is sometimes unappreciated, however, is the fact that model design is not a game-theoretic exercise. Rather, it more properly falls under the purview of the conflict theorist qua conflict theorist. All of which is to say that game theory itself is silent on how the rules of a game are interpreted and represented and on what particular preference and information assumptions are appropriate. Different interpretations of the rules and conflicting preference and information assumptions lead to distinct theories, all of which are subject to formal (i.e., game-theoretic) analysis. In this regard, neither von Neumann nor Schelling has any special expertise. The conclusions they drew remain subject to both logical and empirical scrutiny. And the same is true of any theory of interstate conflict, game-theoretic or otherwise.

For example, there are several theoretically plausible game-theoretic models that are applicable to the relationship of the United States and the Soviet Union during the Cold War (see, *inter alia*, Powell 1990; Fearon 1995; Langlois and Langlois 2006; Sartori 2005; Zagare and Kilgour 2000). Unlike Schelling's explanation of the Cold War peace, it is demonstrable that each of these models is logically consistent. But their theoretical implications are not consistently identical. Thus, the competitive advantage (Lakatos 1970) of one model or another can only be judged empirically, as should von Neumann's purported understanding of the superpower relationship during the Eisenhower administration. Clearly, von Neumann's false prediction about the inevitability of a nuclear conflict in no way reflects on his ability as a game theorist, although in the unlikely case that he actually derived his inaccurate prediction from an explicit game model, then the model would be the culprit, Field's suggestion to the contrary notwithstanding.

References

Brodie, Bernard. 1959. *Strategy in the Missile Age*. Princeton, NJ: Princeton University Press.

Bueno de Mesquita, Bruce, and James D. Morrow. 1999. "Sorting Through the Wealth of Notions." *International Security* 24: 56–73.

Carlson, Lisa J., and Raymond Dacey. 2006. "Sequential Analysis of Deterrence Games with a Declining Status Quo." *Conflict Management and Peace Science* 23: 181–98.

Carlson, Lisa J., and Raymond Dacey. 2014. "The Use of Fear and Anger to Alter Crisis Initiation." *Conflict Management and Peace Science* 31: 168–92.

Danilovic, Vesna. 2002. *When the Stakes Are High: Deterrence and Conflict among Major Powers*. Ann Arbor: University of Michigan Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Fearon, James D. 1995. "Rationalist Explanations of War." *International Organization* 49: 379–414.

Field, Alexander J. 2014. "Schelling, von Neumann, and the Event That Didn't Occur." *Games* 5: 53–89.

Gaddis, John. L. 1986. "The Long Peace: Elements of Stability in the Postwar International System." *International Security* 10: 99–142.

George, Alexander L. 1993. *Forceful Persuasion: Coercive Diplomacy as an Alternative to War*. Washington, DC: United States Institute of Peace Press.

Huth, Paul K. 1999. "Deterrence and International Conflict: Empirical Findings and Theoretical Debates." *Annual Review of Political Science* 2: 61–84.

Jervis, Robert. 1972. "Bargaining and Bargaining Tactics." In *Coercion. Nomos XIV: Yearbook of the American Society for Political and Legal Philosophy*, edited by J. Roland Pennock and John W. Chapman, Chicago: Aldine.

Jervis, Robert. 1979. "Deterrence Theory Revisited." *World Politics* 31: 289–324.

Jervis, Robert. 1988. "The Utility of Nuclear Weapons." *International Security* 13: 80–90.

Johnson, Chalmers. 1997. "Preconception vs. Observation, or the Contributions of Rational Choice Theory and Area Studies to Contemporary Political Science." *PS: Political Science and Politics* 30: 170–74.

Kahn, Herman. 1962. *Thinking about the Unthinkable*. New York: Horizon Press.

Kreps, David M. 1988. *Notes on the Theory of Choice*. Boulder, CO: Westview.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programs." In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, Cambridge: Cambridge University Press.

Langlois, Catherine C., and Jean-Pierre P. Langlois. 2006. "Bargaining and the Failure of Asymmetric Deterrence: Trading off the Risk of War for the Promise of a Better Deal." *Conflict Management and Peace Science* 23: 159–80.

Levy, Jack S. 1997. "Prospect Theory, Rational Choice, and International Relations." *International Studies Quarterly* 41: 87–112.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey*. New York: Wiley.

Martin, Lisa. 1999. "The Contributions of Rational Choice: A Defense of Pluralism." *International Security* 24: 74–83.

McDermott, Rose. 2004. "Prospect Theory in Political Science: Gains and Losses from the First Decade." *Political Psychology* 25 (2): 89–312.

McDonald, John. 1950. *Strategy in Poker, Business and War*. New York: Norton.

McDonald, John, and John W. Tukey. 1949. "Colonel Blotto: A Problem of Military Strategy." *Fortune*, June.

Morgenstern, Oskar. 1961. "Review of The Strategy of Conflict." *Southern Economic Journal* 28: 103–5.

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

Morton, Rebecca B. 1999. *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science*. Cambridge: Cambridge University Press.

Niou, Emerson M. S., and Peter C. Ordeshook. 1999. "The Return of the Luddites." *International Security* 24: 84–96.

Organski, A. F. K., and Jacek Kugler. 1980. *The War Ledger*. Chicago, IL: University of Chicago Press.

Poundstone, William. 1982. *Prisoner's Dilemma*. New York: Anchor Books.

Powell, Robert. 1990. *Nuclear Deterrence Theory: The Search for Credibility*. New York: Cambridge University Press.

Powell, Robert. 1999. "The Modeling Enterprise and Security Studies." *International Security* 24: 97–106.

Quackenbush, Stephen L. 2004. "The Rationality of Rational Choice Theory." *International Interactions* 30 (2): 87–107.

Rapoport, Anatol. 1964. *Strategy and Conscience*. New York: Harper and Row.

Rapoport, Anatol. 1992. "Comments on 'Rationality and Misperceptions in Deterrence Theory.'" *Journal of Theoretical Politics* 4: 479–84.

Rapoport, Anatol, and Melvin J. Guyer. 1966. "A Taxonomy of 2 x 2 Games." *General Systems: Yearbook of the Society for General Systems Research* 11: 203–14.

Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton, NJ: Princeton University Press.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.

Schecter, Stephen, and Herbert Gintis. 2016. *Game Theory in Action: An Introduction to Classical and Evolutionary Models*. Princeton, NJ: Princeton University Press.

Selten, Reinhard. 1975. "A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory* 4: 25–55.

Simon, Herbert A. 1976. "From Substantive to Procedural Rationality." In *Method and Appraisal in Economics*, edited by S. J. Latsis, Cambridge: Cambridge University Press.

Snyder, Glenn H. 1971. "'Prisoner's Dilemma' and 'Chicken' Models in International Politics." *International Studies Quarterly* 15: 66–103.

Von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Walt, Stephen M. 1999. "Rigor or Rigor Mortis? Rational Choice and Security Studies." *International Security* 23: 5–48.

Williams, J. D. 1954. *The Compleat Strategyst*. Santa Monica, CA: Rand.

ZAGARE, FRANK C. 1990. "Rationality and Deterrence." *World Politics* 42: 238–60.

ZAGARE, FRANK C. 1999. "All Mortis, No Rigor." *International Security* 24: 107–14.

ZAGARE, FRANK C. 2004. "Reconciling Rationality with Deterrence: A Re-examination of the Logical Foundations of Deterrence Theory." *Journal of Theoretical Politics* 16: 107–41.

ZAGARE, FRANK C. 2011. *The Games of July: Explaining the Great War*. Ann Arbor: University of Michigan Press.

ZAGARE, FRANK C., AND D. MARC KILGOUR. 2000. *Perfect Deterrence*. Cambridge: Cambridge University Press.

ZAGARE, FRANK C., AND BRANISLAV L. SLANTCHEV. 2012. "Game Theory and Other Modeling Approaches." In *Guide to the Scientific Study of International Processes*, edited by Sara McLaughlin Mitchell, Paul F. Diehl and James D. Morrow, West Sussex, UK: Wiley-Blackwell, 2012, 46–86. Also published online at: http://www.isacompendium.com/.

ZUCKERMAN, SOLLY. 1956. *Scientists and War: The Impact of Science on Military and Civil Affairs*. London: Hamish Hamilton.