

Field, Alexander J.

Article

Schelling, von Neumann, and the event that didn't occur

Games

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Field, Alexander J. (2014) : Schelling, von Neumann, and the event that didn't occur, Games, ISSN 2073-4336, MDPI, Basel, Vol. 5, Iss. 1, pp. 53-89, <https://doi.org/10.3390/g5010053>

This Version is available at:

<https://hdl.handle.net/10419/98505>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/3.0/>

Article

Schelling, von Neumann, and the Event that Didn't Occur

Alexander J. Field

Department of Economics, Santa Clara University, Santa Clara, CA 95053, USA;

E-Mail: afield@scu.edu; Tel.: +1-408-554-4348

Received: 19 August 2013; in revised form: 10 December 2013 / Accepted: 31 January 2014 /

Published: 25 February 2014

Abstract: Thomas Schelling was recognized by the Nobel Prize committee as a pioneer in the application of game theory and rational choice analysis to problems of politics and international relations. However, although he makes frequent references in his writings to this approach, his main explorations and insights depend upon and require acknowledgment of its limitations. One of his principal concerns was how a country could engage in successful deterrence. If the behavioral assumptions that commonly underpin game theory are taken seriously and applied consistently, however, nuclear adversaries are almost certain to engage in devastating conflict, as John von Neumann forcefully asserted. The history of the last half century falsified von Neumann's prediction, and the "event that didn't occur" formed the subject of Schelling's Nobel lecture. The answer to the question "why?" is the central concern of this paper.

Keywords: game theory; deterrence; nuclear strategy; Schelling; von Neumann

1. Introduction

Thomas Schelling is widely thought of, and was recognized by the Nobel Prize committee as a pioneer in the application of game theory and rational choice analysis to problems of politics and international relations. Much of the popularity of his work and other analysis in this vein stemmed from the perception that it contributed to the development and application of new "tools" for understanding and analyzing social phenomena. Following the prize award, the economics journalist David Warsh described him as "the pioneering strategist who made game theory serve everyday economics for thirty years" [1, p. 7].

However, although Schelling makes frequent references in his writings to rational choice and game theory, his analysis of deterrence¹ is based on assumptions about human behavior and logic which, although useful in thinking practically about strategic policy, are at variance with those commonly adduced by game theorists, at least those specializing in its non-cooperative variant.² In areas especially relevant for strategy and conflict, game theory leads to behavioral predictions which are simply not borne out in the laboratory or, as will be apparent, in the real world.

In the Prisoner's Dilemma played once, for example, the Nash prediction is unambiguous: no cooperation. Defection is the strictly dominant strategy. Experimental evidence, however, provides abundant evidence of positive rates of cooperation. Similar "anomalies" are found in voluntary contribution to public goods games (which are multi person Prisoner's Dilemmas), where one sees positive contribution levels, in the trust game, where one sees positive transfers in both directions, and in many other instances.³

Game theory has faced similar predictive failures in its treatment of behavior in the real world. As John von Neumann argued (citations follow), its canonical behavioral assumptions predicted devastating conflict between nuclear adversaries.⁴ This has not happened, and the nonoccurrence of the "most spectacular event of the last half century" was the subject of Schelling's Nobel lecture [2, p. 1]. Schelling could refer to this as an event—something which has taken place—even though it had not—because choice by self-regarding players predicted it so unambiguously. The reality, I will argue, is that because of the disjuncture between human behavior and the self-regarding assumptions often used in formal game theory, the latter offers little guidance, normatively or predictively, in thinking about behavior or strategy in a world of potential conflict.⁵

¹ Deterrence most commonly brings to mind the prevention of attacks on one's own territory or that of close allies. But it can, more aggressively, be used in furtherance of other foreign policy aims. Schelling was interested in both defensive deterrence and its more aggressive forms, and the role that nuclear arms might play in either.

² Non-cooperative theory studies interactions in which players are not allowed to make binding commitments among themselves. Cooperative theory allows such agreements, without specifying or exploring the behavioral attributes that might make them possible.

³ See Kagel and Roth [3], Camerer [4] or Field, [5–8] for more discussion. In the trust game A can anonymously give B some or none of an initial stake, which is multiplied in value in the transfer. B then may, but is not obligated to return as much as she wants to A. If self-regarding players are rational, there are no transfers in either direction.

⁴ One can object that the unitary actor assumption is simply inappropriate when thinking about interactions among states, although one can also object that the approach is inappropriate when applied to individuals (see Thaler and Shefrin [9]). Two points are indisputable: first, von Neumann argued (and believed) that superpower confrontation was a PD, and second, if that was indeed the game, it did not end with the Nash equilibrium. Von Neumann was a pioneer in developing game theory as well as nuclear weapons, and this has resulted in a tension which can be resolved in one of two ways. The first is to argue that nuclear confrontation was not a PD, in other words, that von Neumann *did not know what he was talking about*. The second approach, adopted here, is to accept the PD metaphorically as representative of superpower confrontation, but to argue that the behavioral assumptions that drove von Neumann's (and many other's) thinking were flawed. The central premise of this paper is that the reason we did not and have not experienced nuclear annihilation is that evolutionary history has endowed most humans with predispositions against playing defect in a PD that might well end up being played only once (Field [5,8,10]). People (and states) do indeed sometimes defect. But even when the logic of a strictly dominant strategy is fully understood, *individuals frequently choose not to play it*.

⁵ Developers of formal theory have not been particularly concerned about this, placing more weight on logical consistency and theoretical novelty than on empirical validity.

Before considering in more detail Schelling's evolving acknowledgements of the limitations of game theory in understanding deterrence, it is important to reflect on exactly why the theory is so barren in terms of its implications for policy or behavior. The main reason can be stated simply. So long as agents are self-regarding and there is some possibility of destroying an adversary's offensive capability and/or its will to retaliate, von Neumann was right to characterize nuclear confrontation as a Prisoner's Dilemma.⁶ And, because of the almost unimaginable destructive power of nuclear weapons, particularly thermonuclear weapons, it is a PD that will be played only once if the Nash equilibrium is realized on the first iteration.⁷

In the Prisoner's Dilemma played once, defect (which in this instance means preventive war, preemption, or first strike) is the strictly dominant strategy for both players. As von Neumann argued, it is the only strategy a rational self-regarding player, assuming he is playing against a similar adversary, can choose.⁸ But it is evidently not the strategy chosen by either the United States or the

⁶ A long tradition in the deterrence literature objects, and instead treats nuclear interaction as a game of chicken (see Zagare and Kilgour [11, p. 18]). Chicken involves A threatening to harm B in a way that will also damage A unless B backs off. The best response for either party is to back off in the face of a threat, but if both choose to escalate, the worst (least preferred) outcome ensues for both. Von Neumann did not see nuclear interaction as a game of chicken. Words were cheap. He did not argue that we should try and intimidate the Soviets by *threatening* to attack. He argued for attacking, and for attacking now. Those who reason in this manner tend to downplay or dismiss fears of retaliation, since self-regarding agents would never retaliate ex post (as opposed to threatening to do so ex ante, which would not be credible). A large literature attempts to solve this problem essentially by assuming it away [11, ch. 2].

⁷ Preempters like von Neumann saw the nuclear standoff as a Prisoner's Dilemma in which the strategy space was limited to "attack immediately" or "wait." No truly self-regarding player, reasoned von Neumann, would ever wait. Doing so exposed the actor to avoidable risk and granted a benefit to the adversary (the continued option of preemptive strike). Wait turned the PD into a game of trust: the question then became whether the restraint would be reciprocated. Actual political actors, who are human, often do wait, and establish expectations of reciprocity through diplomacy or other means that sometimes are realized. Von Neumann had little interest in wasting time on the dynamics of deterrence, or why people might make transfers in trust games. His counsel was to launch now, and his argument is unassailable if parties are indeed entirely self-regarding. If one finds this conclusion unpalatable something has to give. This paper argues (and Schelling suggests the same), that deterrence worked and von Neumann's predictions failed because humans are not entirely self-regarding (a position anathema to those who consider themselves hard-headed realists). If that is so, behavioral science will be messier than economists and game theorists might prefer, because the ways in which human predispositions differ from the self-regarding baseline are not deducible from a simple set of first principles. Ideally, the deviations represent empirically validated generalizations from experimental or observational data. That is one of the premises underlying the growing field of behavioral economics, and accepting it means that research methods cannot be limited to blackboard economics.

⁸ Von Neumann assumed that a rational player was by definition self-regarding and he had little tolerance for those who might suggest otherwise. A more general framework for rational choice theory is this: individuals have stable and complete preferences, these preferences are transitive and independent of irrelevant alternatives, and individuals use all available information in choosing action that maximizes a utility function derivable from these preferences. To say that players are both rational and self-regarding is to restrict the range of allowable preferences, and in particular to require that people prefer more material goods to less, and life over death. Suicide bombers may be rational and self-interested, but they are not self-regarding. Neither are those who give all their money to charity because it gives them a warm glow. The more restrictions one places on preferences the more it becomes possible to test the theory against actual behavior. The most general version of rationality, the simple claim that people act in satisfaction of their desires, is

Soviet Union through the four decades of the Cold War. For both sides, defection was trumped by a policy of restraint on first strike coupled with the threat of limited or massive retaliation,⁹ and this was true throughout both the atomic and thermonuclear eras and in spite of substantial shifts over time in the strategic balance between the two adversaries. How and why did this happen, and why did it prevent nuclear war?

From archival sources and interviews conducted by political scientists and historians we know a good deal about discussions that took place in the United States at the highest levels during the first two decades of the nuclear age, and we are learning more about similar debates that took place in the Soviet Union. In the U.S., the central disagreements were between those inclined toward preventive war/preemption/first strike and those recommending policies of deterrence or containment.¹⁰ Support for aggressive preemption was remarkably widespread. It was not limited to a “lunatic” fringe. To provide a compelling rationale for deterrence, one of the objectives of the work Schelling conducted in the 1950s, was to weigh in on one side of a policy debate whose resolution had enormous real world implications.¹¹

Much of what Schelling had to say was based on introspection, casual empiricism, and common sense. To most citizens confronted with the realities of the conflict between the U.S. and the U.S.S.R., a policy of engagement and non-aggressive deterrence seemed intuitively more reasonable than one of aggressive preemption.¹² But nuclear strategists weren’t necessarily like “everyone else”: they prided themselves on asking tough questions, pushing logic to its limits, and, if necessary, thinking about the unthinkable.¹³ The problem faced by Schelling in trying to bolster the case for deterrence with game

unscientific. It is unscientific because it is not possible to validate empirically: there are no observational or experimental data that could disprove it.

⁹ The doctrines of Massive Retaliation and Mutual Assured Destruction differ principally in the proportionality of the response.

¹⁰ These positions are not completely irreconcilable, because one can threaten a nuclear strike to pressure an adversary to do (or not do) something other than simply not launch a nuclear strike against one’s own territory. Indeed, the U.S. relied on such a threat to deter a Warsaw Pact conventional thrust into Western Europe, and some wanted to use the threat to force the Soviets to do other things, such as get out of East Germany, or abandon their atomic weapons. The most aggressive preemption—a surprise attack without prior threats or attempts at bargaining, is nevertheless hard to classify as deterrence. The difference between traditional and aggressive deterrence seems to be captured in the distinction between deterring from and forcing to, although to deter successfully may be to force another not to do something it wants to do—such as attack you.

¹¹ It would be a mistake, however, to think that Schelling, or any of the other strategists or defense intellectuals, adopted a totally consistent position. More often than not they were simply of two minds about a problem, or moved sequentially between positions as they struggled with conundrums that remain with us today. His writings, however, unlike those of Bernard Brodie or William Kaufman, had little direct influence on high level decision making during the Eisenhower administration.

¹² At the start of the Korean War, in July 1950, only 15 percent of the American public agreed that the United States should “declare war on Russia now.” In September 1954 a Gallup poll asked, “Some people say we should go to war against Russia now while we still have the advantage in atomic and hydrogen weapons. Do you agree or disagree?” Only 13 percent agreed. [12, p. 100].

¹³ The reference is to the title of a book published in 1962 by Herman Kahn [13].

theory was that such theory, when coupled with the common assumptions that agents are both rational and self-regarding, provides stronger support for preemption or preventive war.

No one understood this better, or articulated it more forcefully, than John von Neumann, coauthor of the book that helped launch the American intellectual romance with these methods [14]). *Games and Economic Behavior* didn't discuss Prisoner's Dilemmas, which hadn't yet been formally characterized. But von Neumann followed the subsequent literature on non-zero sum games and the equilibrium concept for non-cooperative games developed by John Nash [15]. The Prisoner's Dilemma is a classic venue for applying the Nash solution concept, yet the Nash equilibrium in the one shot PD has long troubled economists because it is so clearly inefficient.

In the PD, you have two plays: you can cooperate or defect. Against the play of cooperate, defect is the superior strategy, and against the play of defect, defect is the superior strategy. Both parties would be better off if they both cooperated, but the logic of strict dominance is unassailable. If the players are rational and self-regarding, we will be hard pressed to explain why such a decision maker should play a strictly dominated strategy.

Von Neumann believed that inasmuch as the US-Soviet standoff was a Prisoner's Dilemma, and inasmuch as both actors were rational and self-regarding, the only defensible policy was immediate attack [12, p. 100]; [16]. Since there was some chance of destroying an adversary's offensive capability and/or will to retaliate by attacking, the best course of action was to launch now. Many others argued in a similar fashion. As the Joint Chiefs of Staff maintained in 1947, "Offense, recognized in the past as the best means of defense, in atomic warfare will be the only general means of defense" [17, p. 77].

One reason the Cold War remained cold was that proponents of preventive or preemptive nuclear war lost arguments in the late 1940s, throughout the 1950s, and again at the time of the Berlin Crisis in 1961 and the Cuban Missile Crisis in 1962.

By the end of the 1960s, most of the dilemmas of the nuclear age remained and if anything had intensified, but key analysts had begun to lose faith in the promise of game theory to illuminate them. The limitations of these methods in providing real guidance to problems of nuclear strategy had become obvious, and doubts could no longer be so easily papered over with optimistic claims that future theoretical progress would remedy these deficiencies [12, p. 261]. After the 1960s strategic studies were less likely to claim to be advancing game theory through the study of nuclear policy, or to be using such theory to untangle such operational challenges as target selection.

The loss of faith in these methods did not mean, of course, that the conundrums vanished.¹⁴ The fundamental policy divide between those inclined to preemption and those inclined toward deterrence remains with us to this day. The world situation has changed since the 1950s, and especially since the early 1990s, with the breakup of the Soviet Union making it no longer as easy to identify adversaries or at least their location. But the heirs to von Neumann's way of thinking continued periodically to occupy prominent positions in the executive branch of the United States Government.¹⁵

¹⁴ Nor did it mean that the elaboration of such models in academic communities ceased.

¹⁵ Setting aside the novel problem of threats from non-state actors, the United States faced a replay of the arguments for preemption/first strike against the Soviet Union in the 1950s as it confronted the prospect of a nuclear armed Iran or

Advocates of preemption always come armed with strong rhetorical advantages. It is a tough minded policy that can appeal to considerations of both opportunism and prudence. To advocate a policy of preemption is to ground policy in the inexorable logic of strict dominance.¹⁶ Rational, self-regarding agents must play a strictly dominant strategy, or, by definition, they are not rational.

The strength of the case for first strike, surprisingly, is, for proponents, not much affected by the military balance between the two adversaries. If a nation is stronger, the argument goes, it must strike first to crush the will, damage command and control, and eliminate as much of the retaliatory capacity of the adversary as it can. And if it is weaker, it must strike first, to benefit from the element of surprise, to use one's assets before they are lost and, by destroying some of one's adversary's offensive capability, mitigate the damage from the almost certain incoming salvo. Why almost certain? Because one's adversary will have made the same calculations, and, whether stronger or weaker, also have concluded that striking first is best. It is a mistake to believe that altering the military balance in either direction will necessarily weaken the calls for preemption from those who favor it.

To advocate a policy of nonaggressive deterrence, in contrast, is to premise policy on human behavior that can be neither recommended nor expected in a world assumed populated by rational self-regarding agents. Deterrence works when both parties play strategies that cannot be defended as strictly rational for self-regarding players, and make inferences that their adversary is not entirely rational or self-regarding. One cannot get beyond von Neumann's case for preemptive war other than by acknowledging this. Advocates of deterrence can, however, point to the principal defect of an aggressive policy of preemption, and it is not a trivial one. As a practical and moral matter, it leads in a conflict between nuclear adversaries more or less directly and more or less certainly to the deaths of hundreds of millions of people.¹⁷

Von Neumann understood that if one wanted to provide a prescriptive justification for first strike, assuming parties are self-regarding, game theory was very effective. Or, if one wanted descriptively to explain why the world had been destroyed in a nuclear conflagration, game theory worked well. Outside of the classroom and world of working papers, in other words, theory could easily succeed at justifying policies that weren't pursued and explaining events that didn't happen. If one wants, in contrast, to justify a policy of deterrence and containment, explain descriptively why a balance of

North Korea, or an unfriendly but nuclear armed Pakistan. And simmering dissension continued after the breakup of the Soviet Union over whether it was prudent to consider nuclear armed Russia as our friend.

¹⁶ A strictly dominant strategy is a strategy that provides a superior payoff, irrespective of the strategy selected by one's counterpart. In the one shot Prisoner's Dilemma, defect is the superior strategy whether one's counterparty cooperates or defects.

¹⁷ This, however, has never been a compelling argument for committed advocates of preemption, since victory is defined as retaining a higher fraction of the surviving population, territory, or economic assets. Obviously, if a country were able to obtain what it wanted from an adversary merely by threatening the use of nuclear weapons, it could be to that country's advantage to do so. But in order for threats to be credible, one must actually be prepared to follow through on them. In the age of conventional war, the victor could often be better off in spite of the costs of fighting. The problem for those pushing the conventionalization of nuclear weaponry, and the idea that one could fight and win a nuclear war, was that by most reasonable standards, this cannot be the case in a nuclear war. As Eisenhower put it, "even assuming that we could emerge from a global war as the acknowledged victor, there would be a destruction in the country (such) that there would be no possibility of our exercising a representative form of government for at least two decades at the minimum" (cited in Jervis [18, p. 62]).

terror kept the peace for forty years during the Cold War, or provide guidance as to what one should do if one does not strike first, formal theory premised on rational choice by self-regarding agents turned out to be of relatively little help.

In the academic and policy worlds, however, those who could claim that they were using game theory to understand real world problems and to provide guidance on how to resolve them enjoyed a premium in the form of career advancement and honorifics. This is a sensitive issue but I think most will agree that it is a fact of intellectual politics that has been true for half a century and remains so today. What then was Schelling's attitude toward formal or pure game theory?

Circa 1960, it can best be described as inconsistent. In the preface to *The Strategy of Conflict* he suggested that he was advancing game theory but later in the book gave mixed signals as to whether he believed such theory, at least at its then current stage of development, could provide practical guidance in matters of nuclear strategy or behavior. The Nobel citation in 2005 nevertheless awarded the prize to Robert Aumann and Schelling for contributions that the committee believed each of them had made to both theory and applications:

The work of two researchers, Robert J. Aumann and Thomas C. Schelling, was essential in developing non-cooperative game theory further and bringing it to bear on major questions in the social sciences. Approaching the subject from different angles—Aumann from mathematics and Schelling from economics—they both perceived that the game-theoretic perspective had the potential to reshape the analysis of human interaction. Perhaps most importantly, Schelling showed that many familiar social interactions could be viewed as non-cooperative games that involve both common and conflicting interests, and Aumann demonstrated that long-run social interaction could be comprehensively analyzed using formal non-cooperative game theory.

...Eventually, and especially over the last twenty-five years, game theory has become a universally accepted tool and language in economics and in many areas of the other social sciences. Current economic analysis of conflict and cooperation builds almost uniformly on the foundations laid by Aumann and Schelling [19].

Much of this was, perhaps by necessity, an exaggeration. Yet it was not entirely accidental that the Nobel committee would suggest, in the press release announcing the prize, and the prize citation, that Schelling, in his 1960 book, “set forth his vision of game theory as a unifying framework for the social sciences.” It is an interpretation that Schelling invited. In the preface to *The Strategy of Conflict*, he advertised the work as “a mixture of “pure” and “applied” research” [20], strictly situating it within the *theory of games* [20, p. v, his italics] and in the text made repeated efforts to incorporate the apparatus of game theory as it was then developed, including extensive discussion and matrix presentations of two person games in normal form.

Schelling's contributions to formal theory are not in the same category as those of John Nash or Reinhard Selten [21]. I don't mean that they are necessarily less or more valuable, simply that they are not in the same category. As Anatol Rapoport, one of the reviewers of Schelling [20] put it,

Dr. Schelling's book is ... not therefore (to) be judged as a contribution to game theory, as a game theorist understands it, but as a contribution to the problem of linking game-theoretical

concepts with other concepts in order to make possible more determinate normative recommendations to the decision maker... [22, p. 434].

There is little in *The Strategy of Conflict* that can be considered an advance in formal theory, and in later decades, Schelling didn't claim otherwise.¹⁸ If, objectively, Schelling did not advance theory, his work nevertheless succeeded in creating an impression in the minds of many non-technical readers—and apparently the Nobel committee—that he had. That he did not push theoretical frontiers should not necessarily be seen as a criticism, since, even after another half century of development, formal theory provides limited insight into the types of problems with which he was most concerned (Walt [23]).

By 1960 Schelling had read Luce and Raiffa [24] in great detail [20, p. vi] and knew that the analytic methods they described were of limited prescriptive or descriptive value in studying the problems with which he (Schelling) was concerned. His acknowledgement of these limitations is to be found partly in Chapter 1, where he allows that models based on the assumption of rational (which he implicitly assumes to mean self-regarding) behavior may be a “caricature” of actual behavior [20, p. 5], but more extensively in Chapter 6:

... some essential part of the study of mixed-motive games is necessarily empirical. This is not to say just that it is an empirical question how people do actually perform in mixed-motive games, especially games too complicated for intellectual mastery. It is a stronger statement: that the principles relevant to successful play, the strategic principles, the propositions of a normative theory, cannot be derived by purely analytical means from a priori considerations ... There is consequently no way that an analyst can reproduce the whole decision process either introspectively or by an axiomatic method. There is no way to build a model ... with the behavior and expectations of those decision units being derived by purely formal deduction ... It is an empirical question whether rational players, either jointly or individually, can actually do better than a purely formal game theory predicts and should consequently ignore the strategic principles produced by such a theory [20, pp. 163–164].

In this and nearby passages, Schelling presages discussions of cognitive modularity (see Barkow *et al.*, [25]), work on dual selves (e.g., Thaler and Shefrin, [9]), and, more generally, the subfield that has come to be known as behavioral economics.¹⁹ And he suggested explicitly that experimental work was strongly needed to mitigate the deficiencies of pure theory: “It does appear that game theory is badly underdeveloped from the experimental side” [20, p. 165]).

An important question is whether subsequent experimental research mitigates the deficiencies of formal theory or simply ends up casting them in a harsher light. Clearly, at the time Schelling wrote *The Strategy of Conflict*, he hoped for the former. He held out the promise of advancing theory so that it could be more useful, writing for example, that “in international strategy the promise of game theory is so far unfulfilled...” [20, p. 10]), suggesting that it would be or could be fulfilled.

Here in clear focus is the conflict between Schelling's ambivalent aspiration to recognition as a theorist, and the acknowledgement that theory, uninformed by behavioral research, and largely

¹⁸ As he said retrospectively regarding his 1960 book, “I don't think I had any noticeable influence on game theorists, but I did reach sociologists, political scientists, and some economists” [26].

¹⁹ “But in the mixed motive game, two or more centers of consciousness are dependent on each other in an essential way” [20].

unconcerned with empirical validation, could not provide the foundation for a science of deterrence. On the one hand the advertisement of the book as a work in both pure and applied research, on the other hand the Chapter 6 recognition of the absolute limits of introspection or axiomatic methods. On the one hand, the acknowledgements that the promise of game theory in this arena was unfulfilled, on the other hand the apparent optimism that it was *so far* unfulfilled.

If one accepts the logic of Schelling's comments in Chapter 6, that one can't reproduce the whole decision process introspectively or through an axiomatic method, that behavior and expectations can't be derived by formal deduction alone, then any hope for advancing a science of deterrence must rest on data: experimental, observational, historical. A science of deterrence had (and has) to be behavioral, *i.e.*, it has to rest on observations of human behavior.

But if progress had to come from the empirical/behavioral side, then Schelling's efforts to push forward theory were investments likely to have low yields in terms of advancing a prescriptive or descriptive science. He acknowledged this (*i.e.*, in Chapter 6) although, as we have seen, not strongly enough or consistently enough to pose obstacles to the Nobel committee's suggestion that he had "develop(ed) non-cooperative theory further" and award a prize based not only on advances in theory but also on its applications to "familiar social interactions".

2. The Evolution of Nuclear Strategy in the 1950s

If we accept Rapoport's judgment that *The Strategy of Conflict* made little or no contribution to formal game theory, we can ask a related question: what influence did Schelling's work have on the practical design of nuclear strategy? His writings provided general intellectual support for deterrence as opposed to preemption, and ultimately to efforts at arms control. At the same time, his emphasis on the manipulation of risk and the prospect of threatening or fighting a limited nuclear war in pursuit of political or strategic objectives can be read as support for an aggressive version of deterrence—brinkmanship—and a potentially dangerous form of saber rattling.²⁰

That said, he does not in fact appear to have had a great deal of influence on the formation of military or strategic policy in the 1950s. Partly this is simply because his government service was mostly in Democratic administrations and that under Truman involved foreign international assistance, not nuclear policy. Even though Schelling is often included in a pantheon of defense intellectuals alongside individuals such as Bernard Brodie, William Kaufman, or Albert Wohlstetter [12, p. 3]), it is actually rather hard to identify his footprints in the history of strategic debates in the 1950s at RAND or elsewhere.²¹ To appreciate this we need to delve more deeply into the actual history of those debates.

²⁰ There is little evidence that either the Soviets or the Americans ever took this counsel to heart: "Rather than being implacable, irrational, or manipulative, states appear to be cautious, flexible, and generally loath to take precipitous action during intense crises" (Zagare and Kilgour [11, p. 228; see also pp. 29–30]; and Zagare [27, p. 113]).

²¹ Citation counts demonstrate that Schelling has had a strong influence on academic thinking about strategic policy, but that is not necessarily the same as having an influence on policy. In this limited influence, Schelling was not alone. Rosenberg describes the impact of strategic thinkers in these terms: "Although such conceptual work was important in shaping public perceptions, and occasionally influenced the thinking of high policymakers or strategic planners, it generally had little relevance in the 1945–1960 period to the pragmatic concerns of operational planners" [28, p. 10].

Proponents of first strike occupied far more than the political fringe, and it is clear that both Presidents Truman and Eisenhower considered seriously the prospect of preventive war or a preemptive nuclear strike against the Soviet Union.²² Neither, however, was prepared to approve an unprovoked surprise nuclear attack against the USSR,²³ and neither was enthusiastic about first use, except, in the case of Eisenhower, in the instance of an actual or imminent conventional attack by Warsaw Pact forces on Western Europe. Each, however, was ready to use nuclear weapons in retaliation for an attack on U.S. territory. U.S. policy was clarified in the Dulles-Eisenhower doctrine (announced in January of 1954) of massive retaliation to either an attack on the United States or a *conventional* Soviet thrust into Western Europe (or possibly other provocations).

The situation in Europe and particularly Berlin created the most practical and immediate concern. The U.S. lacked the conventional forces to thwart a conventional advance in these areas. Air Force war plans, reflecting the doctrine of massive retaliation, anticipated, as a retaliatory response, hitting the Soviets with everything the U.S. had, concentrating on cities. Although, in terms of military planning, Eisenhower was unenthusiastic about preemption except in the case of an imminent conventional attack across Europe, some within the Air Force and the strategic community continued to argue for dispensing with the requirement of a provocation, or maintained that the very fact that the Soviets had nuclear weapons should be considered provocation enough.²⁴ The arrival on scene of the far more powerful thermonuclear weapons did not end support for preventive war, which was widespread among civilian and military elites in the late 1940s and early 1950s.

In 1946 Leslie Groves, the U.S. army officer responsible for shepherding the development of the atomic bomb at Los Alamos during the Second World War, wrote an influential memorandum stating that

If we were ruthlessly realistic, we would not permit any foreign power with which we were not firmly allied, and in which we do not have absolute confidence, to make or possess nuclear weapons. If such a country started to make nuclear weapons we would destroy its capacity to make them before it had progressed far enough to threaten us [12, p. 100].

²² For the evidence, see Trachtenberg ([12, pp. 100–152]).

²³ This prospect was explicitly rejected in NSC-68, which defined U.S. strategic policy in the 1950s, even though the document deliberately exaggerated the Soviet threat (Rhodes [17, p. 106]). NSC-68 was approved by Truman in August of 1950 and declassified in 1977. Truman had of course given the go ahead for dropping two atomic bombs on Japan in 1945. But it was not until 1948 that he allowed the military to proceed with plans for further use of atomic weapons, and he made it clear he was making no commitment that he would use them again. NSC-30, as cited in Jervis [18, p. 24].

²⁴ As Trachtenberg writes: “In the late 1940s, and well into the 1950s, the basic idea that the United States should not just sit back and allow a hostile power like the Soviet Union to acquire a massive nuclear arsenal—that a much more “active” and “positive” policy had to be seriously considered—was surprisingly widespread” [12, p. 100]. As Rosenberg reports, as early as 1947 “the final report of the JCS Evaluation Board on the Bikini tests had recommended that Congress be requested to redefine “acts of aggression” to include ‘the readying of atomic weapons against us’” [27, p. 17]. But, as the public opinion data indicate (see fn. 12), the American public was not enthusiastic about preventive war. As a general rule, humans seem much more prepared to initiate, justify, or approve of attack in the face of provocation than in its absence. Of course, if one can define the mere possession of offensive weapons as provocation, or the possibility that one might acquire such weapons as provocation, the distinction blurs.

Groves' memorandum contained qualifiers about ruthlessness and reasonableness. Others dispensed with qualification, advancing proposals ranging from goal oriented saber rattling (threats) to massive surprise attack. In 1948, William Laurence, the New York Times' science correspondent, recommended an ultimatum to the Soviets: shut down their atomic plants or the U.S. would launch an all-out nuclear war.²⁵ Winston Churchill favored threatening the Soviets: get out of East Germany or the Western powers would destroy their atomic facilities. Leo Szilard pressed for preventive war against the Soviets, simply to wipe out their atomic capability, and at RAND, John Williams was a forceful advocate for similar action. The conservative political thinker (and former Trotskyite) James Burnham was as well. Most remarkably, so was Bertrand Russell, in an address at the New Commonwealth School in London on 20 November 1948. In August of 1950, Secretary of the Navy Francis Mathews gave a speech arguing memorably that the U.S. should become the first "aggressor for peace" [12, p. 117]. Even George Kennan, though he did not endorse initiating nuclear hostilities, mused that a war the Soviet Union stumbled into, before they had a massive arsenal, might be the best solution for the U.S. [12, pp. 103–104].

In the spring of 1953, Eisenhower considered but ultimately rejected the recommendation of a high level study committee headed by retired Air Force General James Doolittle that the Soviet Union be given a two year ultimatum: come to terms or risk global nuclear war. An Air Force study in August 1953 made a similar argument. In May 1954 the Joint Chiefs of Staff Advanced Study Group recommended that the U.S. consider "deliberately precipitating war with the Soviet Union in the near future". Eisenhower would not go along with that, either, and in the fall of 1954, approved an updated National Security Paper which stated (as had NSC-68) that "the United States and its allies must reject the concept of preventive war or acts intended to provoke war" [28, pp. 33–34]. What the declassification of archival information has underlined is not that the U.S. refrained from attacking the Soviet Union (we knew that), but rather how close the country came to doing otherwise, and how forcefully and persistently advocates of preemption pressed their case.

After Dulles's announcement of the doctrine of massive retaliation in 1954, and after it was clear that Eisenhower would not (except in the case of a Warsaw Pact conventional offensive in Europe) approve a preemptive nuclear strike, advocacy of preemption or preventive war became somewhat less open, but enthusiasm for it remained very strong, particularly in the Air Force.²⁶ The option of preemptive war was raised again by three members of the Gaither Committee in the fall of 1957 [28, p. 47]; see also [17] (pp. 106–108). Nor did the consideration of attack die with transition to new political leadership. In the 1960s, Kennedy seriously contemplated a nuclear first strike against the Soviets at the time of the Berlin crisis in 1961, and he was strongly pressured to resort to nuclear weapons in the

²⁵ Laurence was no ordinary journalist. Invited by Groves, he was the only reporter to witness the Trinity blast in New Mexico as well as the atomic bombing of Japan—he interviewed the pilots who flew the aircraft that dropped the bomb on Hiroshima and Laurence himself flew in an observation plane to witness the bombing of Nagasaki.

²⁶ Eisenhower was prepared to launch on warning either of a nuclear strike on the U.S. or a conventional attack on Europe. This can be viewed as preemptive, because attack would be initiated before bombs actually hit the US or Warsaw Pact tanks moved west, but not in the way advocates of preventive war meant it. The distinctions lie in whether the trigger was an immediately impending attack on the U.S. or key allies, the simple possession of nuclear weapons, or the mere possibility that they might be acquired. Prior to the introduction of ICBMs, anticipated warning times for an impending attack were days rather than minutes.

Cuban Missile Crisis in 1962. From a military standpoint, the training requirements for a strike in the presence of a conventional Warsaw Pact build up or in the absence of such direct provocation were essentially the same.²⁷

Nuclear weapons offered the prospect of a relatively cheap alternative to the conventional forces that would otherwise be required to repel a Warsaw Pact offensive westward across the European plain. The difficulty with this strategy was that the Soviets were capable (or at least the U.S. believed they were capable) of promising retaliation in kind. Because the prospect of flattened American cities as a result of retaliation might give American planners pause in the event of a Soviet offensive, U.S. defense intellectuals such as Bernard Brodie and William Kaufman argued early on that the threat to defend Europe in this way was not credible.²⁸

What were the alternatives? Giving NATO countries access to their own nuclear bombs was initially unpalatable to U.S. strategists, so critics of massive retaliation developed a competing doctrine—counterforce—emphasizing a different targeting strategy and more graduated escalation.²⁹ Counterforce entailed responding to a conventional incursion into Europe with a *limited* nuclear response against Russia: hitting airbases, silos, and military installations, but not cities. Perhaps the Soviets would respond against similar targets in the U.S. but, it was argued, they would see their interest in avoiding our cities since we had avoided theirs. We might then threaten to take out their cities one by one until the war could be concluded without it having escalated to an all-out nuclear exchange.³⁰

As an alternative to massive retaliation, the counterforce strategy met with initial resistance from the Air Force, where it was seen, essentially, as soft on Communism (and Communists). In June 1958 Air Force chief Thomas White told an audience of national security specialists he was “disturbed” by the recent tendency “to consider seriously self-restraints in nuclear weapons planning in the face of sure knowledge that no such restraints will be applied by the enemy. Our preoccupation with niceties in nuclear warfare... would, I am sure, delight the Kremlin.” Two years later, however, he supported the strategy. As far as we can tell, this was not the consequence of anything anybody at RAND had written. Why the change?

It had to do with the fact that the Navy’s Polaris program threatened the Air Force’s mission and budget. Submarine launched ballistic missiles (SLBMs), first deployed in 1960, had the great merit of being largely invulnerable, but they lacked the explosive power and accuracy to be used in a counterforce strategy. All they could do well was hit cities [30, p. 244]. Counterforce preserved a role

²⁷ The main difference was how much lead time the military might have in preparing an attack, which would affect how many weapons could be fired off.

²⁸ This reasoning was part of de Gaulle’s rationale for pursuing an independent French nuclear deterrent, targeted at Soviet cities.

²⁹ Today both France and Britain have independent nuclear forces, and as McNamara argued in his 1962 Ann Arbor speeches, such forces make it even less possible to contemplate fighting a limited nuclear war (see Jervis, [18, p. 102]).

³⁰ The ideas of controlled escalation, and competition in risk taking, with the corollary that nuclear conflict might be limited, are significant features of Schelling’s thinking, although they conflict potentially with the viability of a nuclear firebreak, a principle endorsed in Appendix A of *The Strategy of Conflict* and reaffirmed in the Nobel lecture. Moreover, whatever “rules” of limited warfare the Americans wished to play by, the Soviets always ridiculed the notion of using nuclear weapons for bargaining [29, p. 144].

for the more accurate bombers and land based ICBMs³¹ that the Air Force controlled, and was thus a way to marginalize the Navy.

The deterrent value of the Air Force's land based deterrent had been questioned since the early 1950s by Albert Wohlstetter and others because of its alleged vulnerability to preemptive attack by the Soviets, and there had been fierce debates about the merits of hardening the aircraft—putting them in underground hangars—*vs.* dispersing them or simply purchasing more of them to insure that a credible second strike force would survive. Curtis Lemay, head of the Strategic Air Command, brushed aside these concerns. He was confident that the then secret U2 overflights of the Soviet Union would, under any contingency, provide him sufficient warning to get his planes fueled, armed, and in the air. SLBMs promised to solve the vulnerability problem once and for all, but in doing so they threatened the Air Force mission and its budgets.

As a consequence of Polaris, the counterforce approach, which the Air Force but not the Navy would be able to undertake (because of the lower explosive power and poorer accuracy of the Polaris warheads) now had more support within the Air Force. But it was still a hard sell within the Strategic Air Command. When in the winter of 1961 William Kaufman briefed Power, who had succeeded Curtis Lemay as head of SAC, Power angrily responded: “Why do you want to restrain ourselves? ... The whole idea is to kill the bastards... Look, at the end of the war, if there are two Americans and one Russian, we win.” To which Kaufmann replied, “you'd better make sure they're a man and a woman.” Power then reportedly walked out [17, p. 67]; [30, p. 246].

This heated exchange illustrates why the prospect of millions, perhaps tens of millions or hundreds of millions of deaths was, for proponents of preemption, not a compelling objection to first strike. Victory was understood, and continued to be understood, not in absolute but in relative terms: as retaining a higher fraction of surviving population or military/economic assets [18, pp. 59-61].³²

Faced with resistance in SAC, and lack of enthusiasm from Eisenhower, who feared, given the dynamics of what he would later call the military industrial complex, that moving to flexible response was an invitation to launching aggressive war, Air Force chief White ultimately gave a weak endorsement to the counterforce strategy. Early in the Kennedy administration, however, these ideas came into ascendance, along with the idea that conventional forces should be built up to avoid having to choose between the unpalatable consequences of responding to an attack with massive retaliation and the equally unpalatable alternative of doing nothing. But counterforce, which on the face of it seemed highly attractive to those appalled by the likely consequence of either first strike or massive retaliation, brought with it its own set of issues.

³¹ Technical change during the 1950s was making land and air based weapons smaller and more accurate.

³² Carl Kaysen, in the 1961 first strike plan formulated at the height of the Berlin crisis, formalized this criterion somewhat less colloquially: “Accompanying these assumptions is the notion that prevailing in a general war means coming out relatively ahead of the enemy. As an example, if the US has lost 20% of its industrial capacity and 30% of its people, but the Sino-Soviet bloc has lost 40% of its industrial capacity and 60% of its people, then the US, somehow or other, has won the war” [31, p. 13].

3. Targeting Coordination

In the late 1940s the U.S. Navy actually condemned nuclear weapons as “immoral” [30, pp. 232–233]. This changed when the service, along with the Army, had the opportunity to begin acquiring its own tactical nuclear weapons. By the end of the 1950s, with the multiplication of nuclear armaments under the control of three of the four armed services (only the Marines didn’t have them), the obvious need for more coordinated targeting generated pressure for a single integrated operational plan (SIOP).³³ The SIOP for the fiscal year 1962 (effective 1 April 1961) that emerged from the Eisenhower administration and was inherited by President Kennedy and Secretary of Defense McNamara anticipated, as had earlier war plans, a massive preemptive nuclear strike against the Soviet Union, Eastern Europe, and China in the event of an actual or even *impending* Soviet conventional attack on Europe.³⁴

SIOP-62 represented the most aggressive posture advocates of first strike could obtain in the absence of a President willing to approve preemption (in cases other than an impending conventional attack across Europe). The overriding imperative in this plan was to provide a first strike capability in the event of conventional incursion in Europe, and to insure that a massive second strike capability would survive any attack by the Soviets on the U.S. and, in the event the U.S. was attacked, actually be used. Military planners saw plan rigidity as a merit, not a defect. SIOP-62 did provide a range of options in terms of how many missiles would be fired; this depended on the amount of advance warning or preparation time that the Air Force would be given.

For all of the options, the intent was to kill as many Communists as possible as quickly as possible. With a one hour warning, the retaliatory plan anticipated firing the 1459 U.S. nuclear weapons kept on alert (these ranged from 10 kilotons to 23 megatons, for a total of 2.164 gigatons); conservative estimates were that 175 million Russians and Chinese would die. If the President gave the go ahead for a full preemptive strike (28 h advance notice required) all 3,423 weapons with a total of 7.847 gigatons would be launched, and 285 million Russians and Chinese would die. These casualty estimates did not include deaths in Eastern Europe or victims of fallout around the world, and they reckoned damage only from blast, neglecting destruction resulting from heat, fire, and radiation. A more comprehensive estimate placed likely casualties closer to 1 billion [17, p. 88]. The option for a full preemptive strike was labeled Plan 1-A, giving insight into the priorities of some of the war planners ([28, p. 6]; [30, pp. 269–272]; [31]).³⁵ Total American megatonnage reached its peak in 1960, although total number of warheads peaked in 1966 [17, pp. 89, 95].

³³ Each Air Force Command controlled the detailed plans for the use of its weapons, as did the Army and the Navy with its tactical weapons. With the Navy acquiring strategic weapons (Polaris), the situation was worsening, with little coordination and much duplication of targeting [30].

³⁴ In an earlier briefing on SIOP-62, before Kennedy took over, David Shoup, Commandant of the Marine Corps, asked SAC chief Power what would happen if the Chinese were not involved in the fighting. “Do we have any option so that we don’t have to hit China?” Power responded, “well, yeah, we *could* do that, but I hope nobody thinks of it because it would really screw up the plan” [30, p. 270].

³⁵ Kaplan had access to critical documents related to SIOP before they were reclassified under the Reagan administration (some have since been declassified a second time).

SIOP-62 mirrored the inflexibility of strategy which had been a feature of its predecessors, the Joint Outline Emergency War Plan, which evolved in the 1950s into the Joint Strategic Capabilities Plan, the Joint Mid-Range War Plan, and the Joint Strategic Objectives Plan, all of which were characterized by plans for massive retaliation or attack [28]. These plans had called for executing a massive retaliatory nuclear strike in the event of a “General War,” defined in the Joint Strategic Capabilities Plan as “an armed conflict in which Armed Forces of the U.S.S.R. and those of the United States are overtly and directly engaged.” The Army tried in 1958 to have the following words added: “as principal protagonists with the national survival of both deemed at issue” but the Air Force succeeded in having the amendment nixed [30, p. 277]. As a practical matter, the battle between preempters and deterrers was to define how sensitive would be the trip wire that would trigger massive retaliation. Provocation would still be required for the U.S. to launch, but the trigger would not have to be very substantial—a shooting incident in Berlin would have been enough—and the plans were so inflexible that they were in essence a Doomsday machine. Defenders—and one could find some support for this view in Schelling—argued that it was this inflexibility that made it such an effective deterrent.

McNamara and Kennedy initially wanted a wider range of options—the possibility of more flexible response. The successor operational plan, SIOP-63, reflected some degree of counterforce thinking. It divided Soviet targets into five separate categories, with the initial U.S. strike only on such strategic sites as air and missile bases and submarine pens. Under Eisenhower, the desideratum of inflexibility was reflected in the facts that Minutemen missiles had to be fired in groups of 50 or not at all and they were each rigidly preprogrammed to strike one target. The Air Force initially refused to reprogram its missiles to provide flexible targeting, acquiescing only after its funding had been cut for a month. McNamara’s deputy pointed out that if the military chiefs really believed in SIOP-62, there was little need for generals [30, pp. 280–281]. In some sense this was a rules *vs.* discretion battle; the generals worried that any flexibility weakened the deterrent power of nuclear weapons and possibly the resolve of civilian leaders to use them if necessary.³⁶

But as McNamara succeeded in replacing massive retaliation with somewhat more flexible response (although even the limited attack options in SIOP-63 involved massive megatonnage with huge civilian casualties from fallout), some of the defects of counterforce became more apparent. In August of 1960 the newly orbited KH-1 (Corona) satellites began generating photographs indicating that the missile gap, which Kennedy had successfully exploited in his Presidential campaign, like the previously touted bomber gap, was an illusion.³⁷ Eisenhower had been right in denying the existence

³⁶ The objection to flexibility is that it becomes an end in itself, a way of kicking the can down the road. With policy makers unable to decide in advance what they would do in various contingencies, the imperative became to preserve flexibility so that decision makers would be faced in the heat of crisis or battle with a choice among options they had not been able or willing to make with minds unpressured by immediate circumstance (see [18, p. 80]). The more flexible response is built into war plans, the greater the potential strain on communication and control, and the greater the likelihood that there will be no response, one reason the military tended to be averse to increased flexibility.

³⁷ The claim of a missile gap originated with the Gaither report in 1957, a panel set up by Eisenhower that included Paul Nitze, a Democrat, and as Richard Rhodes puts it, “went rogue” in terms of what Eisenhower had expected it to do. Nitze was in fact the main author of the report, according to McGeorge Bundy [17, pp. 106–108]. Rereading this history is a reminder that the manipulation of evidence that Iraq possessed WMDs in 2003 did not represent the first time U.S. intelligence estimates and have been influenced by political and bureaucratic imperatives.

of such a gap, at least in favor of the Soviets. In fact, as the satellite reconnaissance data revealed, the U.S. had a 10 to 1 advantage in ICBMs. The Soviet strategic forces were so small and disorganized that counterforce now began to look appealing as a first strike option that might succeed in wiping out or very significantly degrading the Soviet retaliatory capability. The September 1961 National Intelligence Estimate concluded that the Soviets had just four operational ICBMs.

In 1961 Khrushchev was threatening a separate peace treaty with the East Germans and in August began to build the Berlin Wall. The American garrison in Berlin had enough food, fuel, and ammunition to survive without resupply for just 18 days. So, with very limited options if the Russians further tightened the screws in Berlin, Kennedy considered a preemptive nuclear attack on hardened Soviet missiles as an option. The plan, little known to this day, was developed by Carl Kaysen, who estimated that only five to thirteen million Americans might die [30, pp. 297–298]. The Soviet military capability would have been devastated, but they would still probably have been able to hit back at the United States with a few megatons, and “New York and Chicago, with their great concentrations of people, can be virtually wiped out by a small number of high yield weapons. In thermonuclear warfare,” Kaysen added, in case it was not apparent, “people are easy to kill” [31].

Counterforce, which had originally been proposed as a move away from the hair trigger (and possibly non-credible) policy of massive retaliation, now provided additional fuel for preempters as well as grounds for more nuclear weaponry: more powerful, more accurate, and more of them. If the Soviets were ahead of us, argued the Air Force, we needed more missiles and bombers to deal with our vulnerability. If they were behind us, as turned out to be the case, we needed more weapons to transform counterforce into a viable first strike capability and to stay ahead of them.

However—as a practical matter—as the U.S. continued to become strategically stronger, it could be argued that a counterforce targeting strategy increased the likelihood that the Soviets would launch a preemptive strike against the U.S.³⁸ Increased Soviet vulnerability, argued some, including Schelling, could paradoxically increase the threat to the U.S. If counterforce was destabilizing in this way, it would be better to go back to targeting cities, holding them, essentially as hostages.

McNamara, who initially endorsed counterforce because it offered the prospect of fighting a limited rather than an all-out nuclear war, now found that his success in championing it provided justification for even greater demands on the part of the Air Force for weaponry, which he wanted to restrain as he built up conventional forces. Although SIOP-63 was never altered to reflect this, McNamara soon cooled on counterforce, and began emphasizing the importance of Assured Destruction, eventually Mutually Assured Destruction. It was enough, he and Kennedy decided, if the U.S., having absorbed a Soviet first strike, could destroy a quarter of the Soviet population and half its industrial capacity. In 1962, Kennedy and McNamara decided unilaterally to limit US land based ICBMs to 1054 (1000 Minutemen plus the existing 54 Titans), down from the 10,000 pressed for by SAC and the 3000 ultimately requested by the Air Force [17, p. 95].

³⁸ This was particularly so in the early 1980s when President Reagan committed to a missile defense system. Although most are aware of how close the U.S. and U.S.S.R. came to nuclear war in the Cuban Missile Crisis, the two countries came extremely close again in 1983 because the Soviets, listening to the bellicose rhetoric of the Reagan administration, and observing the unprecedented U.S. peacetime military buildup, became convinced the U.S. was preparing for a first strike against them (see [17], [29, pp. 345–346]).

By 1965 McNamara had come full circle, arriving at something close to the Eisenhower/Dulles doctrine of massive retaliation that had prevailed a decade earlier.³⁹ McNamara's disenchantment with counterforce and return to 1950s era strategic doctrine coincides roughly with what Trachtenberg [12] has identified as the beginning of the exhaustion of strategic thinking.⁴⁰

For the last half century we have, in a sense, been replaying old tapes. None of the issues involving how to make a threat credible, whether to target cities or military assets (counterforce or countervalue), whether we should prepare for assured destruction or limited conventional war, whether or not flexible response is desirable—none of these issues is new. All were identified and actively debated prior to the mid-1960s. The arguments raised in the 1970s and 1980s, ranging from those articulated by members of the Committee on the Present Danger (established in 1976) to the pressure to establish an “independent” Team B challenge to the CIA's estimates of Soviet capabilities and intentions (1976), to those that almost led to war with the Soviet Union in 1983, were the same as had been articulated earlier, with in many cases the same cast of characters or their protégés [17, pp. 124–126, 150–157]. It is symptomatic of the exhaustion of strategic thinking—and a tacit acknowledgment that game theory premised on rational self-regarding agents could not ultimately offer much help in formulating strategy—that after 1966, Schelling moved away from issues of nuclear policy (see, e.g., [32]).

Schelling's contribution to targeting debates during the heyday of strategic thinking is unclear, in part because he was often of two minds about many of the issues. He spent a year at RAND in 1959 and was thus connected to and familiar with the think tank out of which the ideas of graduated escalation and counterforce emerged. But Schelling was ultimately lukewarm toward these concepts. Certainly people like William Kaufman were more central in articulating and advancing the doctrine. Schelling ran war games at Camp David in September 1961, with Blue and Red Teams consisting of U.S. military strategists, including among others John McNaughton, Alain Enthoven, Carl Kaysen, McGeorge Bundy, and Henry Kissinger. The results, which might be considered an unusual form of behavioral research, revealed rather striking inhibitions against going nuclear, even in a small way [30, p. 302].

³⁹ US policy cycled back again to emphasize counterforce in the 1970s. Following Vietnam, Schelling moved away from direct involvement with the government and military policy. In 1970 he led a faculty delegation protesting Nixon's involvement in Cambodia; this action effectively ended his role as a defense intellectual. See Schelling [26].

⁴⁰ “Strategy as an intellectual discipline came alive in the United States in the 1950s. A very distinctive, influential and conceptually powerful body of thought emerged. But by 1966 or so, this intellectual tradition had more or less run its course [12, p. 261]... there was an intellectual vacuum in the whole national security area. The economists, and people heavily influenced by their style of thinking were for a variety of reasons drawn into this vacuum. What they had was something very general, a way of approaching issues, rather than anything that in itself suggested substantive answers that went right to the heart of the strategic problem. Looking back at this body of thought as a whole, it is clear that the publication of Schelling's *Arms and Influence* in 1966 marked something of a climax. After 1966 the field went into a period of decline: the well seemed to have run dry, the ideas were by and large no longer fresh or exciting...” [12, p. 44].

I would go beyond this, and, to extend the metaphor, suggest that there never was water in the well: The methods were as barren of useful insights in the 1950s as they were acknowledged to be by the mid-1960s.

Subsequently Schelling directed an interdepartmental group within the National Security Council which “examined certain long-range aspects of political military planning.”⁴¹ The report, *A Study of the Management and Termination of War with the Soviet Union*, completed a week before Kennedy was assassinated, looked at how the U.S. and Soviet Union might bargain to bring a nuclear war to conclusion under a variety of different scenarios [33].

Schelling’s writings, which emphasized controlled escalation and limited shots across the bow to advance national objectives but also stressed trying to keep war from getting out of control, were more consonant with SIOP-63, with its flexible response, than SIOP-62, with its inflexible massive retaliation. But Schelling was ambivalent about counterforce: Trachtenberg describes Schelling in 1960 as “being pulled in both directions”, indicating that he ultimately came down in favor of a controlled counter-population (countervalue) strategy, which some of his colleagues found cruel or bizarre [12, pp. 35–38]. Certainly counterforce allowed for graduated escalation, which Schelling favored, whereas if massive retaliation were to be viewed as entailing bargaining it was going to be a pretty short conversation.

Counterforce, on the other hand, risked blurring the bright line between conventional and nuclear weapons, which as emphasized in his Nobel lecture and Appendix A of *The Strategy of Conflict*, he valued. And one could argue, consistent with Schelling’s writings, that the inflexibility of massive retaliation/cities only/mutual assured destruction made the policy a more effective deterrent and thus, arguably, contributed to preventing war. The more inflexible the response, however, the greater the danger from false alarms (false indications of incoming missiles or bombers). The merits of counterforce were nevertheless ambiguous and Schelling’s attitude towards it conflicted, as indeed McNamara’s came to be. Schelling’s support for a bright line between conventional and nuclear weapons was at odds with the elements of his thinking that emphasized bargaining, competition in taking dangerous risks, flexible response, and graduated escalation.

Schelling’s overarching framework, like that of Clausewitz, emphasized that war and diplomacy should be considered elements of a broad spectrum of bargaining behavior, and the emphasis on risk manipulation can be seen as support for a particular type of brinkmanship [12, p. 45].⁴² The scenarios worked through in the 1963 NSC report were illustrations of how this might work. Of course, neither Schelling nor anyone else has yet had actual experience fighting and bargaining within the context of a nuclear war. Alain Enthoven once shut down a general who questioned Enthoven’s expertise by noting that he had fought just as many nuclear wars as had the general [30, p. 254]. No one really knows how or whether one could bargain in a controlled way in the heat of threatening or actually exchanging salvos with a nuclear adversary. There is an enormous range of problems. For example, with civilian and military targets often only a few miles apart, how could one have counted on the Soviets to understand, in the face of a barrage of incoming missiles, that the attack was counterforce only, and thus be persuaded not to go after U.S. cities?

⁴¹ Under the Kennedy administration, Schelling participated in a number of interagency committees, including one that led to the establishment of the hotline between Moscow and Washington. See Schelling, [26].

⁴² There is some irony here given the general antipathy of Democratic policy advisors to what were sometimes seen as the dangerous and reckless policies of Dulles and Eisenhower.

Although Schelling could explore the conduct of nuclear war only through simulations, he did have an opportunity to apply his insights to the waging of conventional war. The experience was not a happy one. He was asked in 1961 by Paul Nitze, then Assistant Secretary of Defense for International Security Affairs, to come to Washington as his arms control deputy. Schelling demurred, but recommended his friend John McNaughton, with whom he had worked in the early 1950s administering the Marshall Plan. McNaughton at first resisted, saying he knew little about arms and strategy, but Schelling promised to teach him everything he needed to know. McNaughton went to Washington, where he was instrumental in persuading the Pentagon not to block the limited Arms Control Treaty of 1963, which banned atmospheric testing. When Nitze became Secretary of the Navy in 1963, McNaughton became Assistant Secretary of Defense.

In 1964, he was charged by McGeorge Bundy with developing an “integrated political-military plan for action against North Vietnam.” The plan, based on Schelling’s ideas about how to wage limited war, was to use graduated escalation, employing large numbers of troops as a deterrent to the North’s invasion of the South and then applying pressure on the North through an air campaign. The idea was to wait for, or possibly invite a provocation from the North and then retaliate with a measured air campaign to force North Vietnam to change its behavior.

In planning that campaign, McNaughton visited Schelling to try and figure out the answers to several questions: what did the U.S. want North Vietnam to do or stop doing, how would bombing make them alter their behavior, how would the U.S. know this had happened, and what would prevent the North from reverting to what it had been doing previously once the bombing stopped? The two strategists were unable to come up with answers to any of these questions, although Schelling did advise McNaughton to limit the bombing to three weeks.

The campaign, code-named Rolling Thunder, failed to alter the behavior of the North Vietnamese, if anything, hardening their attitude, and solidifying their will [30, pp. 334–335].⁴³

4. Taking Data to the Theory

Thomas Hobbes’ work *Leviathan* (1651, [34]) has for decades been understood as the classic non-formal evocation of the Prisoner’s Dilemma. Schelling does not place the PD front and center in his 1960 book (there are a few references to it at the end) although it is central to the problems he explored. This centrality is recognized in the Nobel citation, which began with an evocation of the Hobbesian dilemma:

Wars and other conflicts are among the main sources of human misery. A minimum of cooperation is a prerequisite for a prosperous society. Life in an anarchic “state of nature” with its struggle of every man against every man is, in Thomas Hobbes’ (1651) famous phrase, “solitary, poor, nasty, brutish, and short”.

⁴³ The problem with using threats or pressure effectively is that it requires that the entity threatened respond rationally. In contrast, successful deterrence requires a number of areas in which irrational logic and thought processes must prevail (refusing to attack in the first place; actually retaliating after deterrence has failed). The hubris of much strategic thinking comes in its confidence that one can know or specify in advance in what realms rational thought processes will and will not prevail, both in oneself and in one’s opponent.

The citation goes on to provide a capsule history of the development of game theory. But it is hard to see, realistically, how advances in pure theory have helped us, or are likely to help us in understanding the behavior of nuclear adversaries. One can read Schelling's position on game theory in *The Strategy of Conflict* as inconsistent, or characterize it as acknowledging its limitations but suggesting that with improvements it could do the job.⁴⁴ In his review of Schelling [20], Rapoport [22]) questioned the likelihood of this. He argued that the challenge was not simply to improve upon a slightly flawed approach:

These ideas, especially some of the striking paradoxes, are interesting and stimulating. I believe, however, that they indicate the necessity of transcending game theoretical thinking (i.e., thinking exclusively in strategic terms) rather than the need to incorporate into the theory of games matters which do not fit into its conceptual repertoire... The fact remains that there is no rationally justifiable conclusion that leads the two players of a Prisoner's Dilemma game without communication to insure for themselves the largest joint pay-off. Such an outcome can result only if "irrational" considerations are allowed to determine the choice of strategy, for example, "solidarity," "trust," "the determination to do the right thing, no matter what the consequences may be," etc. Such considerations have until now been anathema to the realists. Among the strategists, it is perfectly proper to advocate "calculated risks" based on bluff, blackmail, and intimidation, but risks based on trust (which admittedly may be misplaced, else the risks would not be risks) fall automatically outside the scope of strategy because the associated concepts are not even in the vocabulary of the strategist.

Rapoport is not always completely on target (it is now well understood, for instance, that the PD with communication is formally identical to the game without it) but his basic point is well taken. Schelling respected Rapoport—there is a footnote in *The Strategy of Conflict* referring to one of his earlier essays as magnificent [20, p. 7], but we may infer that this review struck too close to home. When Schelling reviewed Rapoport's *Strategy and Conscience* three years later, his tone was uncharacteristically harsh [35]. Rapoport's book is a sprawling affair, containing at its best some deep insights into probability theory, and some lucid critiques of the strategic way of thinking and its dangers, but one can legitimately criticize it from a number of angles. Nevertheless, one is struck in reading Schelling's review by the animus he appears to bear towards the author. Schelling appears quite angry, angry that Rapoport has tarred all strategists with the same brush, has failed to distinguish good theory from bad, and didn't have the courtesy to contact his targets to ask them what they in fact thought (rather than, perhaps defensibly, inferring it from their writings). My interpretation is that at the time Schelling felt compelled to defend those whose aim was to apply and improve game theory, a group in which he included himself. And yet, if one focuses on Chapter 6 of Schelling [20], Rapoport's and Schelling's understandings of and acknowledgements of the limitations of formal theory were not very far apart.

⁴⁴ Still, if his position was that the theory lacked much predictive or explanatory power now, but might, with improvement, have it in the future, what was the justification for extensive inclusion of game theoretic apparatus in a book devoted to analyzing pressing current problems?

The referenced passages from *The Strategy of Conflict* and Schelling's review of Rapoport are the basis for the interpretation that Schelling was inconsistent in acknowledging or addressing the shortcomings of game theory as a prescriptive and descriptive guide to behavior, and conflicted about its prospects and his role as a theorist. These ambivalences, if anything, became deeper with time. In a symposium in his honor at Harvard University in October 2006 [36], Schelling observed that the Nobel citation gave him cause to reflect on whether he "was then or ever had been a game theorist." He recalled that he learned what theory he knows from Luce and Raiffa [24], a book he spent "more time studying than the Holy Bible," but noted that the Nobel committee seemed most impressed by work he had published before Luce and Raiffa appeared, and before he had read it.

This last is a very puzzling claim, since the Nobel citation states that Schelling received the prize for "developing non-cooperative game theory further", and it is hard to see how he could have made progress on this account prior to digesting Luce and Raiffa. Certainly, the avidity with which (by self-report) he consumed the work⁴⁵ suggests that he had high hopes it would be useful to him. By the end of 1950s, however, as evidenced in *The Strategy of Conflict*, he was aware of its limitations. That realization would have been unsettling to anyone schooled in the merits of rational choice analysis, and doubly troubling in a world in which rewards (honorifics, professional advancement) for apparently applying game theory to real world problems remained so high. In Schelling [20], any concern on this account was papered over by taking the optimistic position that yes, the theory was limited in its predictive or explanatory value, but future progress would resolve these deficiencies.

Formal game theory can be beautiful from an aesthetic perspective and challenging in terms of the mathematical and logical puzzles it presents. Social scientists benefit from a basic understanding of its concepts and principles if only to appreciate its limitations and because it has become part of the intellectual landscape. But if one is interested in solving real world problems, or using theory to understand human behavior, it is quite often a cul de sac. Ariel Rubinstein has probably been most forthright about this, arguing that formal theory simply should not be "a tool for predicting or describing real human behavior" [37, p. 616].⁴⁶ Rubinstein did not, however, necessarily conclude from this that research in game theory be shut down. A close reading is that his plea is that experimentalists and students of the real world stop harassing theorists with the disjuncture between theory and behavior.⁴⁷ Howard Raiffa (one of the coauthors of Schelling's "bible"), reached a somewhat similar conclusion, largely abandoning the theory program in his later career, and instead focusing on the practicalities of negotiation rather than its mathematical analysis (compare [24] with [38]). Anatol Rapoport said essentially the same thing as Rubinstein forty years earlier:

The theory of games has been developed much beyond the zero-sum game, and it is not the fault of the theoreticians that the results are so frequently indeterminate or psychologically disturbing. The mathematical theory of games was never meant to be a behavioral theory, but

⁴⁵ He recalls that he spent between 100 and 200 hours reading it (Schelling, [26]).

⁴⁶ For evidence of similar attitudes in political science, in particular the assertion that logical consistency is more important than empirical validity, see Walt [39]).

⁴⁷ Economists often talk of taking theory (or a new model) to the data. The objection here seems to be to the efforts of behavioralists and experimentalists to take data to the theory.

only a mathematical one, which examines the internal logic of certain situations without necessarily drawing conclusions about what this internal logic may imply in human affairs [40, p. 437].

There can scarcely be any “problem in the social sciences” of greater import than that of understanding relations among nuclear adversaries. It is not surprising if what theory has actually delivered in this area has been a disappointment if the theory was never meant to be behavioral.

5. The Logic of First Strike

If agents are rational and self-regarding, there can be no logic of deterrence, only one of first strike. Von Neumann argued famously, with respect to the Soviets, “If you say why not bomb them tomorrow, I say why not today? If you say today at 5 o’clock, I say why not 1 o’clock?” ([12, p. 104]; Blair, 1957, cited in [16, p. 143]). As he neared death in 1957, he confided to his friend Hilary Putnam, a philosophy professor at Harvard, about his pessimistic view of the future. Von Neumann told Putnam he was “absolutely certain (1) that there would be a nuclear war and (2) that everyone would die from it” [41, p. 114].⁴⁸

Von Neumann’s analysis and conclusions were common at the time among intellectuals and those who had thought seriously about nuclear war, both before and after the arrival of thermonuclear weapons. Three years later, on the front page of the New York Times, the British Novelist C. P. Snow made a similar prediction, stating that absent massive disarmament, thermonuclear war within a decade was a “mathematical certainty.” As Schelling pointed out in his Nobel lecture, nobody at the time found this claim exaggerated [2]. Given what we know of the Emergency War Plan of the late 1950s or SIOP-62, or the Berlin crisis of 1961, or the Cuban Missile Crisis of 1962, or U.S.—Soviet tensions in 1983, this is not surprising. Von Neumann, had he still been alive in 1960, might have asked, why not sooner?

The failure accurately to predict has not inhered in the logic. The problem is with the implied behavioral assumptions. A taboo not codified in any treaty has kept nuclear weapons from being used against an adversary since 1945, just as a taboo kept poison gas from being used in the Second World War. In his Nobel lecture Schelling is rightly interested in these normative constraints on behavior and the degree to which they give rise to behavior at odds with what was predicted and counseled by von Neumann. Some behavioral inhibitions—in my view as much biological in origin as they are cultural—checked the pressures on or inclinations of leaders to attack. Our survival has depended upon the fact that decisions makers were and are, in some respects, the opposite of rational and self-regarding. It has depended in part on the fact that they were and are human.

There is no practical defense against ICBMs. There is therefore, as Richard Rhodes puts it, “no military solution to safety in the nuclear age” [17, p. 101]. Game theoretic analysis, which commonly assumes that players are logical, rational, and self-regarding, leads to the conclusion that the surest and

⁴⁸ Von Neumann hated communists and had no moral qualms about working on the H Bomb project. From an early date he looked forward to nuclear conflict between the two superpowers. He expressed this anticipation in a 1951 letter to Lewis Strauss: “I think that the USA-USSR conflict will very probably lead to an armed ‘total’ collision, and that a maximum rate of armament is therefore imperative” [30, p. 63].

most effective way to reduce this vulnerability is to launch a surprise attack aimed at an adversary's offensive weapons. Because players are human, however, and sometimes prone to retaliate to attack if they can, even if it gains them little or nothing, or even makes their deteriorated situation even worse, this is an almost certain invitation to mutual incineration. The effort to ground strategy in game theory produces, using Robert Jervis' words, "doctrines which are incoherent and filled with contradictions, which though they are superficially alluring, upon close examination do not make sense" [18, p. 19].⁴⁹

In his Nobel autobiography, Schelling tried to interpret the convention against use of nuclear weapons as an example of a focal point. But this is not an appropriate use of the concept. Pure games of coordination have multiple equilibria, and the concept of a focal point refers to the role of tacit or implicit knowledge in enabling players to coordinate on one of them. The Prisoner's Dilemma is not a pure game of coordination. It has only one equilibrium (unlike the New York meeting problem that Schelling popularized), and that equilibrium is inefficient, unlike any of the equilibria in a typical coordination game. The no-nukes "convention" is not a Nash equilibrium, as disgruntled advocates of preemption have pointed out again and again, and there are strong incentives, pressures, and temptations for players to deviate from it, in contrast to what is true for a coordination game equilibrium once reached. If there are norms (whether biological or cultural in their origin) which incline us to the cooperative solution in PDs, they are different from the social or cultural norms that contribute to a focal point. The former require in some respects that we be the opposite of self-regarding, whereas the latter (such as those, within a common language area, that allow us to agree on the meaning of a word) do not (Field, [42]). We are best served by acknowledging these differences rather than suggesting that these types of norms are of the same genus.

The analysis of games of coordination does help us understand the role of tacit or implicit knowledge in solving such problems as which side of the moving walkway to stand on, or where, in a pre-cell phone age, and in the absence of prearrangement, we should meet in New York. In that sense, Schelling has contributed to the use of game theory to understand "everyday economics", to use Warsh's words, or "familiar everyday problems", to use the words of the Nobel citation.

But the economic and social significance of coordination problems pales in comparison to those presented by Hobbesian dilemmas, and the Nobel citation's emphasis on Schelling's contribution to the analysis of "non-cooperative games that involve both common and conflicting interests" makes it clear that they had in mind the latter category of social dilemmas. The reality is first that Schelling made little contribution to the formal development of non-cooperative game theory and second that such theory is of little value in understanding how humans solve Hobbesian, or Prisoner's Dilemmas. On both of these counts, then, the Nobel citation reflects a certain amount of misdirection.

In the long run it matters little why someone wins the Nobel Prize. What does matter in the social sciences is correctly identifying what types of contribution advance our understanding of human behavior. The challenge of understanding what contributes to collective action or collective inaction

⁴⁹ He goes on to argue, "To argue that any nuclear doctrine must be at least partly irrational does not mean that all doctrines are equally at odds with the reality they try to reflect and shape. If one starts with misleading conceptions, the more complete and thorough the reasoning, the stranger and more confusing the results. Only by understanding and accepting the implications of nuclear weapons can we develop a more appropriate policy. But even such a policy cannot meet all the standards we normally require of rationality" [18, p. 20].

(restraints on harm) beyond the family are central to social and behavioral science. Claims to advance must be carefully vetted, with the most important criterion not formal axiomatic consistency but the extent to which actual human behavior is explained or predicted. Misdirection with respect to the contributions of game theory has been facilitated by a finding of common ground among those who love the intellectual challenges it poses, even as they acknowledge its limited applicability to actual human behavior (e.g., Rubinstein), and a broader audience that understands enough about the approach to appreciate its apparent potential, but not enough to recognize its limitations.

Deterrence can only work if in some respect humans are not rational in their logic and/or self-regarding in their preferences. A strategy of defensive deterrence has two main pillars: (1) each counterparty refrains from first strike; and (2) will in fact retaliate in the event deterrence fails. Neither pillar, von Neumann appreciated, could be defended as the behavior of a self-regarding rational actor. To von Neumann, failure to strike first was irrational and foolish. It was foolish because it exposed one's country to the risk of being hit first. And since the nuclear counterparty had access to similar logic, and assuming the adversary was also a self-regarding rational actor, the attack must be imminent.⁵⁰ If missiles were not already inbound it was only because the adversary was also behaving in a foolish manner. To refrain from attacking first was to put one's faith in irrational behavior on the part of one's counterpart. And, if one's counterpart were behaving in a foolish fashion, failure to strike first meant passing up an opportunity to gain at her expense. In a thermonuclear world, to refrain from first strike was to refrain from defecting in a Prisoner's Dilemma game which would have more than one play only if both parties irrationally chose not to defect.

Some have explained the failure of the United States to wage preventive war in the early 1950s as due to its self-perception as relatively weak [12, pp. 100–115]. Yet in 1961, when intelligence revealed the disarray and weakness of Soviet strategic forces, and Khrushchev's nuclear bluster had been shown to be largely a bluff, the U.S. again passed on the preemption option. Conditions were more favorable for a U.S. first strike than they had been at any time since the late 1940s. And in his September 1961 war games, Schelling could not get either the Blue or the Red teams to go nuclear, no matter how hard he provoked. In the policy domain, the 1961 Kaysen plan at the height of the Berlin crisis was rejected, even as a contingency [30, pp. 299–300]. Something obviously restrained the willingness of even tough minded strategists like Kaysen and Kissinger to act on the logic of first strike. Von Neumann must have turned over in his grave.

As many have pointed out, the second pillar of deterrence, the promise of retaliation, also requires behavior that cannot be justified as rational. Having decided (irrationally) to refrain from first strike, the threat of retaliation is intended to deter a counterparty from doing the rational thing, which is to attack first. But if the attack has already happened, deterrence has failed. With many U.S. cities in ruins, of what possible value would have been the destruction of millions of Soviet lives, simply to prove, after the fact, that we had "meant" what we said when we promised/threatened retaliation?

The threat of massive retaliation would not be credible if the counterparty believed one were completely rational because it would make no sense to carry through on it after the fact. This train of logic leads to the conclusion that a rational actor should not and would not be deterred from attacking

⁵⁰ See Kahn, ([13], pp. 151–152, cited in Rapoport, [40], p. 134). Kahn describes an imagined conversation between a Soviet general and Khrushchev as they consider the pros and cons of a first strike on the United States.

by a threat of retaliation so long as she assumed the target was rational, *because a rational victim of aggression would not retaliate*. It is easy to threaten retaliation, but talk is cheap. This reality helps illuminate the obsession of military planners with inflexibility, as they have struggled over the years to make retaliatory threats credible.

Schelling argued that madmen could not easily be deterred by threats of retaliation [20, p. 6]. An apparent corollary is that those who are rational (not mad) *can* be deterred. But as this analysis reveals, you cannot deter a rational actor who also believes you are rational, because your threats of retaliation will not be credible.

This issue exposes one of the soft underbellies of strategic thinking, and is one that Schelling chose not to probe too deeply. In *Arms and Influence* [43] he cited Max Lerner, writing in *The Age of Overkill* (1962) that “The operation of the deterrence principle in preventing war depends upon an almost flawless rationality on both sides.” ([44, p. 27], cited in Schelling, [43], p. 229). Without directly challenging or contradicting Lerner’s claim Schelling dryly notes that “...when people say that “irrationality” spoils deterrence they mean—or ought to mean—only particular brands of it” [43, p. 229].

The practical success of deterrence rests in part on the fact that the very act of refraining from first strike, which is demonstrably *not rational*, may increase the credibility of a threat of retaliation. By contravening the counsel of von Neumann and other aggressive preempters, an actor illustrates by her restrained behavior a willingness to behave in an irrational although still apparently instrumental fashion. Thus a counterparty might, for this reason, think twice about dismissing the threat of massive retaliation on the grounds that such retaliation would, after the fact, be irrational.

Strategists like von Neumann argued that a nuclear war initiated by a U.S. first strike could be won, if the U.S. possessed sufficient strategic superiority, and this argument surfaced again and again in the 1960s, 1970s, and 1980s. There were several targeting strategies to choose from. One could focus principally on military targets, to degrade or eliminate second strike capability. Or one could focus on CCC—command, control and communication—decapitation strikes that killed leaders and destroyed communications capability, so that even if weapons survived and decision makers were (irrationally) angry and prepared to retaliate, the weapons would not be fired. Or one could stick with the traditional cities only strategy, demoralizing the Soviets (“Shock and Awe”) to make sure that there was little or no (irrational) will or desire left for retaliation. Advocates of preemption stressed that in order for the U.S. to face a risk of retaliation, both the Soviet capability and the will to retaliate had to survive our first strike. And even if there was some probability that both did survive, winning the war could be defined as the U.S. suffering lower relative losses.

Success for the U.S. as aggressor would require combining an assumption of rational first strike on our part with irrational restraint on first strike on the part of the Soviets (otherwise the Soviets would already have attacked and U.S. would not have been able to strike first) followed by rational behavior by the Soviets following our strike (why bother retaliating since their effort to deter us had obviously failed). If, in contrast, we believed the Soviets would be consistently irrational in their behavior (irrational restraint combined with irrational retaliation in the event we attacked) we might, if we were rational, be deterred from attacking.

But note that the non-aggression outcome associated with symmetric adoption of mutual assured destruction required asymmetric assumptions about the rationality of each of the adversaries.

In particular, each had to believe that they were rational but their adversary was not—which is objectively impossible.⁵¹

To navigate in these clouded waters we will be less helped by game theory than by an understanding of social psychology, the theory of mind,⁵² and such regularities as the fundamental attribution error.⁵³ Contrary to Lerner’s argument, successful deterrence requires not flawless rationality but elements of irrationality—sometimes real, sometimes perceived—on the part of both parties. It is likely that the predispositions that allow deterrence to work have in part a biological foundation (see Field, [5–8,10]).

If the scenarios for aggressive war pushed by von Neumann and others seemed to gamble with the nation’s fate (what if the assumptions were off by a bit; was it OK if “only” five to thirteen million Americans died?), preempters had a ready answer. Their proposals were driven not just by opportunistic calculations, but by fears that the Soviets were making exactly the same calculations, planning to fight and win a limited nuclear war (see, e.g., Kaysen [31], or arguments advanced by the Committee on the Present Danger in the 1970s and 1980s). And if that were true, it was clearly in the U.S. interest to attack first. Would those who hesitated prefer confronting the aftermath of a Soviet first strike on the United States, with degraded hardware, disrupted command and communication, and/or devastated cities? From this it’s a short distance to von Neumann’s if you say 5 PM, I say why not 1 PM. The U.S. military was acutely aware of vulnerabilities in its own command and control systems, and resisted modifications in SIOP-62 on this account. It wanted inflexible response. The concern was that even if U.S. weapons survived a Soviet strike, the relevant individuals⁵⁴ might not be able or willing to launch them.

Schelling, of course, preferred to win without fighting, or by fighting as little as possible (“a theory of deterrence would be in effect a theory of the skillful nonuse of military forces” [20, p. 9]). But he acknowledged at least implicitly the asymmetric but complementary roles rationality and irrationality would have to play in successful deterrence, although they were the reverse of what was needed for successful first strike.

Schelling repeatedly emphasized the potential role played by irrationality:

⁵¹ Robert Jervis was exploring similar issues when he wrote, in 1984, that “A rational strategy for employing nuclear weapons is a contradiction in terms” [18, p. 19].

⁵² Theory of mind involves such questions as how animals (including humans) infer another’s knowledge or intentions from the direction of its gaze, the timbre of its voice, or the expression on its face (for discussion, see Cheney and Seyfarth [45]). In game theory, none of this matters: there is promise of a parsimonious short cut to inferring intention. Sometimes, but by no means always, this short cut provides good predictions, but where it fails, and when the setting is nuclear confrontation, the failures are particularly problematic.

⁵³ The error, which has been widely demonstrated, is to assume that my behavior is governed by the situation while yours is governed by your disposition, in short, to assume that I’m rational and you’re not. Since the success of mutual assured destruction requires mutually asymmetric assumptions about the rationality of the two parties, deterrence could not succeed in the absence of this cognitive bias. The classic study is Ross [45]).

⁵⁴ It would be perhaps a misnomer to call them decision makers, because the whole point of inflexible response was that they were not supposed to be making decisions.

“It is not a universal advantage in situations of conflict to be inalienably and manifestly rational in decision and motivation. Many of the attributes of rationality, as in several illustrations mentioned earlier, are strategic disabilities in certain conflict situations” [20, p. 18].

But he does not confront the theoretical incoherence of a model that requires that each adversary believe that they are rational whereas their adversary is not, or that they are rational but must make their adversary believe the opposite, or that they are or will be rational in some instances but not others. We have already seen that a rational attacker (A) would not find credible the threat of retaliation by B if A attributed rationality and self-regarding preferences to B. A cannot be deterred by B if she attributes rationality and self-regarding preferences to B. If A is rational and self-regarding and has not attacked, it must be because of her attribution of some degree of irrationality to B (in other words, A must have some fear of retaliation). But if that is so, the merely *threatened* use of nuclear arms by A in an attempt to alter B’s behavior in ways distinct from simply refraining from attacking would also lack credibility. If A is deterred today by B’s irrational threat of retaliation, why should that be any different tomorrow? Aggressive preemption or the threat of aggressive preemption by A will succeed only if B combines irrational restraint prior to A’s attack (else why would A have the opportunity to strike first?) with rational calculation of B’s self-interest following A’s strike.

If we are wedded to a rational choice modeling approach, should we not expect at least a consistency in the irrationality of B before and after A’s rational strike? Experimental research makes it abundantly clear that humans are inclined to have their behavior influenced by both rational and irrational thought processes, and it sometimes makes a big difference which prevail. *Formal game theory offers us no help in estimating those probabilities*. If your fate and the fate of the world depended upon it, would you rather your political leaders be good intuitive psychologists, skilled at inferring the emotions, motives, beliefs, and intentions of counterparties, and of themselves, or well-trained game theorists?

To accept the argument that nonaggressive deterrence (peaceful coexistence) requires a combination of irrational and rational behavior on the part of both parties is not to reject such strategies as worth pursuing, nor is it to argue against being deterred in the face of a determined and committed opponent. Such policies and behavior often are worth advocating and pursuing, and as an empirical matter often have worked as they were intended. The problem for game theory is that in a world of self-regarding agents, there is always a stronger (more hard-headed, more tough-minded) case to be made for preemption. Allowing oneself to become too absorbed in the strategic way of thinking runs the risk of finding oneself in a race to the bottom, in which aggressive policies systematically trump those that are less so.

Leslie Groves’ advocacy of a policy of destroying the relevant facilities in any country that might threaten to develop nuclear weapons sounded much like Vice President Cheney’s advocacy of a broad license for preemption more than half a century later: “If there’s a 1 percent chance that Pakistani scientists are helping al-Qaeda build or develop a nuclear weapon, we have to treat it as a certainty in terms of our response. It’s not about our analysis, or finding a preponderance of the evidence. ...It’s about our response” [47, p. 62]. Since a 1 percent probability is low, its estimation subjective, and one Cheney proposed the United States arrogate to itself, there was little limit, if this doctrine were accepted, to the range of targets an attack upon which could still be clothed in the rhetoric of defense.

What is troubling about a policy of preemptive or preventive war, as applied to Iraq, and possibly to Iran or North Korea, are the implications for the U.S. if it is adopted by other nations. What it means is that any nation, on its own say so, would be justified in attacking the United States or any other country if it judged that there was even a very small probability that the United States might attack it. From a standpoint of rational choice theory and a model of a world consisting of self-regarding agents, there is no error in this reasoning.⁵⁵ International law has always recognized the right of a nation to protect itself by attacking an adversary when an attack from that adversary is “imminent”.

The Cold War spanned four decades, and its epicenter can be said to have been Berlin. The conflict began in 1947, two years before the blockade and airlift. Conflict over the city brought us close to nuclear war in 1961. The Cold War ended with the fall of the Wall in 1989 and the subsequent collapse of the Soviet Union (1991). For U.S. baby boomers the Cold War was an omnipresent fact of life, a feature of it from their birth until, to the surprise of many and consternation of some, the conflict suddenly ended a quarter century ago. The Cold War defined the political, strategic, and intellectual milieu within which Schelling’s most influential work was conducted.

Over most of that period, a philosophy and strategy of containment and deterrence helped prevent thermonuclear war between two well-armed adversaries, the Soviet Union and the United States. Neither launched a surprise attack on the other. The question I have asked in this essay is whether the success of the Cold War stalemate depended upon the rational and self-regarding behavior of both parties, and on each party’s belief that the other was rational and self-regarding. If it did, then the interaction of these two adversaries would indeed be a fitting subject for game theory. Lerner claimed that it did and it was. Schelling danced around the question. To challenge the assumption made by Lerner that successful deterrence required consistent (flawless) rationality on the part of all parties too directly might have undercut the impression that game theory/rational choice approaches were central to the conclusions of his analysis.

That said, Schelling appears to have recognized the flaws in Lerner’s claim. Without acknowledging the poor predictive power of models premised on rational choice by self-regarding agents, in particular the fact that individuals often do refrain from first strike, even when rational analysis indicated, as von Neumann protested repeatedly, that it is in their interest to do otherwise, and that they often will respond in kind to attack, even when this makes no sense *ex post*, there would have been no explanatory space for the topics that interested Schelling. There would be no arena for the murky world of threats and promises, driven by conflicting behavioral predispositions within individual humans themselves and characterized by the asymmetrical ability to benefit others by failing to do them harm as opposed to providing them with affirmative assistance. It is his acknowledgment of that world of shadows, in which things were not always what they seem, and what

⁵⁵ Note that a similar issue applies to the abandonment of the Geneva conventions, which former Attorney General Alberto Gonzalez characterized as “quaint.” There is little reason from a game theoretic perspective why the Geneva conventions should be respected, but clearly to some degree they are and have been, and many in the U.S. military objected to the US adoption of methods of torture, which resulted in the deaths of tens of captives, on the grounds that such actions threatened the protections available to captured American soldiers.

seemed to be the pursuit of flawless rationality was sometimes nothing of the sort that makes Schelling's work interesting.⁵⁶

6. Rational Choice

The Royal Swedish Academy of Sciences saw it differently. Rather than interpreting Schelling as having shown that behavior interpreted as the pursuit of flawless rationality was often nothing of the sort, the Academy's prize committee claimed that Schelling (and Aumann's) contribution was to show the opposite: "A consequence of these endeavors is that the concept of rationality now has a wider interpretation; behavior which used to be classified as irrational has become understandable and rational" [19, p. 3]. This is misdirection. It reflects the same kind of papering over of contradictions which renders most discussions of U.S. strategic policy incoherent. Human behavior is what it is, and sophistry cannot make the play of a strictly dominated strategy rational. If x is black, and we understand what black means, one can't make it white simply by reinterpreting it as white.

In describing behavior, the word rational is often used in different ways. The broadest meaning is simply the claim that people act in satisfaction of their own desires. This version is not interesting from a scientific standpoint, since it is impossible to conceive of any behavioral data that could not be made consistent with it. A narrower version posits that people have goals reflected in preferences, that these preferences are both stable and transitive (if A is preferred to B, and B to C, then A is preferred to C), and that people use all available information to choose a course of action likely to lead to the realization of these goals.

The third and most rhetorically powerful version of rational choice carries forward all of the language about goal seeking, but adds that preferences are, in addition to being stable and transitive, also self-regarding. Following Gintis [48], I suggest that self-regarding is superior terminology to self-interested, since it avoids the possible ambiguity arising when individuals experience a "warm glow" from helping others. Such helping behavior (and I do not mean to suggest that all helping behavior has this characteristic) might arguably be self-interested, but is not what we would mean by self-regarding.

In using the term rational in this paper, I have meant this third, most restrictive, and most rhetorically powerful use of the term. Under this meaning, to assume rational choice is to assume that people (or countries) act so as efficiently to advance their material self-interest.

⁵⁶ Schelling's work is as much about diplomacy as the art of war fighting (as witness the title of his 1966 book: *Arms and Influence*). But, as those skeptical of diplomacy have always argued, talk is cheap, and actions speak louder than words. If these precepts are taken to heart, diplomacy is not part of the strategy space. As von Neumann understood, so long as players are rational and self-regarding, the strategy space was limited to launch or not launch, and the case for the former was unassailable. The idea that threats or promises might be used to further political or strategic aims presupposed that actors would behave in ways that could not be defended as rational. A country might threaten to retaliate, or threaten to build a Doomsday machine to overcome the prospect of weakness of will, but a rational actor would never do either. The idea, popularized by Schelling, that, short of attacking, one might create a situation where things might get out of control and an attack "might happen" would have struck von Neumann as mealy mouthed. "Just do it" would have been his reaction.

7. The Logic of First Strike—Once More with Feeling

It is a commonplace among tough minded thinkers to say that talk is cheap, and actions speak louder than words. But rhetoric matters. It affects human attitudes and behavior, even if theory often suggests it shouldn't. Most humans are inclined to support military operations in response to what is seen as an unprovoked attack. Public opinion tends to be less comfortable with preemptive attack, because the threat countered is probabilistic and thus speculative. Aside from the small fraction of humans with sociopathic tendencies, attack simply for personal or territorial aggrandizements is the toughest sell of all: most humans are repelled by it. Even the most cynical dictators find it necessary and desirable to justify aggression in terms of a prior litany of perceived wrongs. For this reason Hitler felt compelled to cloak his invasion of Poland in 1939 as preemption to counter what he claimed was an intolerable threat from that country. The immediate *casus belli* was a ginned up raid on Germany supposedly conducted by Polish forces.

In 1949 there was not a 1 percent probability that the Soviets were obtaining the bomb: there was a 100 percent chance that they had it. But surprisingly, this is beside the point in terms of the argument for attacking now. The logic of first strike is inexorable for an analyst consistently applying "realist" assumptions. It does not depend much on the size of this probability. After all, one could imagine a hypothetical conversation between von Neumann and former Vice President Cheney, "If you say attack if there's a 1 percent probability, I say why not if there's a 0.1 percent probability. And if you say 0.1, I say why not 0.01." Because of the asymmetrical ability to benefit others by not harming them, as opposed to providing affirmative assistance,⁵⁷ an asymmetry which is grossly magnified in the case of weapons of mass destruction, one requires very low probabilities in order to make the case for preventive war on these grounds.

Now let us again consider again the second pillar of MAD: the promise or threat of retaliation. In order for such a threat to play any role in deterrence, it must be credible. This is, of course, a persistent theme in Schelling's writings, and much subsequent game theoretic analysis.⁵⁸ But let us put ourselves in the position of a submarine commander somewhere in the Pacific at the height of the Cold War. A devastating first strike has been launched by the Soviets which has wiped out New York, Chicago, and Los Angeles. With 16 MIRVed missiles in his tubes, the commander receives the order to launch a retaliatory strike on Soviet cities, one which, in 20 to 30 min, will kill thirty million people.

Does it make any sense to extract the launch codes, and proceed with firing? What will be accomplished by killing an additional thirty million people? Reputation is irrelevant because this is not a game that will continue. So the submarine commander, if he is rational, cannot logically justify setting in motion the launch procedure, and a rational President cannot logically justify ordering his submarine commander to do so.⁵⁹ But it gets worse.

Since game theory generally presumes that our counterparty is as rational and logical as we are, the U.S. forecasts the likely response of the Soviets to a first strike attack, and concludes that their promise

⁵⁷ "One of the lamentable principles of human productivity is that it is easier to destroy than create" (Schelling, [43, p. v]).

⁵⁸ At the October 2006 Kennedy School event, Schelling identified one of his principal concerns as "How can you make a promise that's believable when it's clear that left to your own devices you'd rather not do it" [23].

⁵⁹ Some of his advisors felt that President Reagan had doubts he would be able to issue the launch order in such a circumstance, one of the reasons he was so strongly attached to the idea of a missile shield (see Rhodes, [17]).

of massive retaliation is hollow—cheap talk. Since if they are rational, they will not retaliate, we can afford to launch first strike against them with impunity.⁶⁰ Moreover, since they will have realized that if the U.S. is rational, our threat of massive retaliation is also hollow, they will have concluded that they can launch with impunity, if we are foolish enough not to have already launched. In each case, the appeal of attack lies both in the prospect of removing a sword of Damocles and in the possibility of territorial aggrandizement or increased world influence. In the absence of any effective defense, there is an enormous game theoretic advantage to the offense, to moving first (Lee, [49, p. 198]).

Schelling understood that no realistic study of international relations, just as no serious study of human behavior, could be premised on the assumption that agents are or should be motivated solely by self-regarding preferences. It is true that such preferences are strong, and underlie what I have called the foraging algorithms (Field, [7]). They cause us to seek food when we are hungry, and water when we are thirsty. But we also possess behavioral predispositions specialized for social interaction, including some that bias us in favor of refraining from first strike, some that make us willing to engage in costly (and thus other-regarding) punishment, even of third party violators, and a heightened sensitivity to detecting those who cheat on the social norms that commonly reflect these species typical behavioral predispositions.

Schelling doesn't address the evolutionary pathways that might have created this. I and others have, and make a strong argument that one must posit the operation of selection at levels higher than the individual organism (group or multilevel selection) in order to allow this to have happened (Wilson and Sober, [50]; Boehm, [51]; Field, [5, 6, 8, 10]; Wilson and Wilson, [52]).

Once one allows for this restraint, once one acknowledges a species typical inhibition on first strike, one which cannot be defended as rational, and one which would not have been favored upon first appearance by organism level selection, one has the foundation for a realistic science of human behavior. One has a framework within which it becomes possible to understand why von Neumann's prediction has, so far, been proved wrong. It becomes possible to see why policies of deterrence and containment can appeal to "natural" human impulses, as much if not more so than the press toward preemption and first strike, which draws support from the logical, rational capabilities associated with the prefrontal cortex. Whereas the counsel of our predispositions specialized for social interaction often agrees with that proffered by our foraging algorithms, sometimes they conflict, and sometimes the former trump or short circuit the latter.

⁶⁰ Thus Powell does not appreciate that in assuming that "no political objective is worth certain destruction", the problem of the credibility of US nuclear retaliation to a conventional Soviet attack on Western Europe is the same as the credibility of the threat of nuclear retaliation to a Soviet attack on US territory. And in stating what many would view as self-evident, that with secure second strike, "there is no situation in which it is rational for a state deliberately to launch a nuclear attack first" [53, p. 15]. Powell is simply wrong: this conclusion depends on assuming that one's adversary will in fact retaliate. I am not saying one *should* attack first; if one is rational one will do so only if one believes one's opponent will rationally not retaliate. Irrational human propensities to retaliate are part of the reason in the real world that attacks are deterred. For a similar reason, Jervis is wrong to state "There would be no reason for the Russians to hold back once Americans had destroyed what they value most..." [18, p. 74]. There would be no reason if they were human, but lots of reason if they were rational—since such retaliation would gain them precisely nothing so long as gains are defined in terms of material interest.

Of course the strength of these inhibitions varies among individuals, for reasons both biological and related to personal history. Of course a skilled leader can defeat them, by demonizing the enemy, or by conjuring threats, both real or imagined. But the point is that the inhibitions are real, they are species typical, and they do have to be defeated (the task of doing so is a central part of military training). It is because of these inhibitions, combined with the real and ever present possibility that they can be defeated by appeal to prudence or rational self-interest, that we have the world we live in, a dangerous world, but one which has been sufficiently peaceful to allow the human population to increase to over 7 billion.

Without these irrational inhibitions, the shadowy world of threats and promises, of the “uglier, more negative, less civilized part of diplomacy” (Schelling, [43], p. vi), would simply not exist. There would be no space for it because conflict would never be looming over us. It always would already have started. But it is the world as it exists, peopled with individuals who have irrational as well as rational behavioral predispositions, that Schelling wished to analyze and understand. That world is one in which the promise or threat of harm becomes a critical element of the grammar of social intercourse, and a central concern of Schelling was how people or nations could be influenced, how these threats or promises are or could be effective. Our world, and the fragile peace in which we live most of the time, presupposes non-Nash behavior. It is a world in which human agents, much but not all of the time, refrain from playing defect even though logic counsels us that defect is the only defensible play.

For most game theorists today, the modeling crucible within which to study the emergence and maintenance of cooperation (non-defection) is the indefinitely repeated game.⁶¹ This is theoretically convenient, because equilibria in which people cooperate (as well as those in which they don't) can, within this context, be attributed to rational self-regarding choice. But it is empirically and historically a poor choice if one is concerned with interactions among agents who have a power to harm each other which is asymmetrically larger than their ability to help each other (aside from failing to harm them). In thinking about behavior among adversaries armed with hydrogen bombs it is absolutely essential that we explain how a Prisoner's Dilemma that might well end up being played only once is successfully surmounted, in the sense that neither party defects. To do so one must account for why players are prepared to choose a strategy which in theoretical terms is strictly dominated. Any realistic study of human behavior, whether at the individual, small group, or country level, must begin with the acknowledgment that humans possess some behavioral predispositions that cannot be defended as the rational behavior of a self-regarding agent.

If one or both parties defects in the real time Prisoner's Dilemma, there is no point discussing the logic of deterrence: there is simply no arena for it. Once one enters a state in which a fragile peace is sustained through mutual (and non-rational) restraint on first strike, successful deterrence requires persuading an adversary that you have the intention, under certain states of the world, to behave in a manner that could not at that point be defended as rational. The behavior, and the propensity of humans to indulge in it, is captured well in experiments involving the ultimatum game [54].

The world Schelling (and we) inhabit is a shadowy place where what appears intuitively to make perfect sense cannot in fact be defended as the behavior of a rational agent, where we can talk

⁶¹ Robert Aumann's contribution to their analysis was a major theme in the announcement of his Nobel Prize. Aumann shared the award with Schelling in 2005.

seriously, if not coherently or consistently, about why it may be advantageous to appear to be irrational. It is a world of conundrums and distorted mirrors, in which we are likely to be startled by our own reflection.

No gas was used by combatants during the Second World War, and to a remarkable degree, the Geneva accords on the treatment of POWs were adhered to by combatants, even though the U.S. never signed these conventions. No nuclear weapons were used in the Korean War, and the U.S. did not bomb across the Yalu river. The Chinese/North Koreans did not bomb bases in Japan, American ships at sea, or bases in Japan (Schelling, [43], pp. 129–130), and Chinese bombers never departed directly from China, always effecting a wheels down in North Korea before pursuing their targets (Schelling, [2]). Thus, the fury on the battlefield notwithstanding, even war has historically been fought *with restraint*, albeit with different amounts of it in different conflicts.

But these restraints, and others like them, always chafe. And within elements of countries such as the United States, there remained and remains a chafing, a chafing at the constraints apparently imposed by a defensive, deterrence based policy, a chafing at the constraints imposed by international diplomacy as symbolized by the United Nations and the various “threats” of world government, a bugaboo that Robert Welch of the John Birch Society railed against. With the breakup of the Soviet Empire and the end of the Cold War, and with the victory of President Bush in the 2000 election, and finally with the unprovoked attack on United States soil on 11 September 2001, voices calling for aggressive preemption were once again front and center, as they had been periodically in the past. In foreign policy, it is true that Iraq was an obsession, but the larger agenda was to redefine our strategic posture.

The Pentagon was to become part of a department of Offense as much as Defense, and the one percent doctrine advanced by Vice President Cheney gave the United States extraordinary scope for launching military action where and when it saw fit. Restraint on first strike was to be drastically attenuated: attacks according to this doctrine could be justified by the merest threat (“one percent probability”) of possible attacks on the United States, a threat that would be evaluated and defined by the United States. And along with the attacks on the presumption against the launch of offensive war came attacks on the quaintness of the Geneva conventions, opening the way to the use of interrogation techniques the U.S. had branded as torture in war crime prosecutions after the Second World War.

As a practical matter the country moved in directions periodically advocated by people like James Burnham, Curtis Lemay, Herman Kahn, John von Neumann, and Barry Goldwater. Science fiction writers and Hollywood screenwriters had fantasized about worlds in which individuals would be incarcerated before they committed crimes, simply because statistical methods predicted a high probability that they would commit offenses. The one percent doctrine represented the application of the principles reflected in the 2002 movie *Minority Report* to countries. As Suskind put it, “Where once a discernible act of aggression against America or its national interest was the threshold for a U.S. military response, now even proof of a threat is too constraining a standard” [47, p. 214]. Although the influence of preempters receded with the evidence of the lack of WMDs in Iraq and the election of Barack Obama, von Neumann’s heirs will surely regain their seat at the policy making table again in the future.

Schelling’s work on deterrence was premised on the assumption that the United States would not itself launch offensive war: that our task rather was to create conditions where we could safeguard our

security in a world in which others might. A world of deterrence and containment lacks the simplicity, clarity and dreadful beauty of the Nash equilibrium in the Prisoner's dilemma. It is a messy world, cluttered with paradoxes, in which arguments can be made on almost all sides of any policy recommendation. But it is the world we must live in if we are to avoid Armageddon. Schelling's work was not about "thinking the unthinkable" to borrow title words from Herman Kahn's 1962 book. It was, in part, about avoiding the unthinkable. And to avoid the unthinkable we must have a sound, empirically based picture of the human ethogram, one which acknowledges the sometimes conflicting behavioral predispositions with which we are endowed. People emphasize that we live today in a world of WMDs against which there are only limited defenses. The challenges today are different than they were in the Cold War, but they are not entirely novel. In spite of the hundreds of billions of dollars spent on defense the nation stood completely defenseless against an attack of Soviet ICBMs until the end of the Cold War.

The reality is that in some of the most consequential types of human interaction, formal game theory has not been useful for understanding how people behave, or how they necessarily should behave. Where it is clearly wrong in its predictions, it can, however, serve a useful purpose in helping us break out of the box in which much of modern social science has imprisoned behavioral science. Its usefulness in this fashion becomes apparent when it makes clear, unambiguous predictions which are not borne out by data. In a number of important instances, such as the Prisoner's Dilemma played once (a game commonly dismissed as "uninteresting" or "too restrictive" by game theorists), its predictions are abundantly contradicted by experimental and observational evidence. But this reality can have less impact than perhaps it should, because some theorists simply aren't interested in using the theory as a tool for understanding human behavior (Rubinstein, [37]). Deep down, many theorists wish to develop it as a logically consistent and internally coherent set of analyses unconstrained by a requirement that its predictions actually map onto human behavior. Such a posture is not reconcilable with a serious commitment to an empirically based social or behavioral science.

Because of its abstruse, esoteric quality and apparent rigor, many, including economic journalists, continue to be attracted and intrigued by game theory and, more generally rational choice models, without fully appreciating their limitations. Over his long career Schelling did little to discourage those who saw him as a "pioneering strategist" making game theory "serve everyday economics". But the substance of his work has been premised on acknowledgements of a more complex human ethogram than game theory has been able to accommodate.

If we combine the assertions that deterrence often does work with the argument that it presupposes a commingling of rational and irrational logic and thought processes, then we are forced to question the dominant behavioral theory and motivational assumptions which are thought to underpin realist foreign policy. There are certain conclusions in modern social science which, although indisputably correct, are considered bad form to bring up. A prime example is the absence of an instrumental political rationale for voting in national elections. A second is the inability to defend or explain deterrence when agents are rational and self-regarding. Deterrence works because we are human, not because we are entirely rational. Both of these conclusions point to the limitations of game theoretic approaches. Schelling understood these limitations (even if he did not advertise them) and so should we.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Warsh, D. *Knowledge and the Wealth of Nations: A Story of Economic Discovery*; Norton: New York, NY, USA, 2006.
2. Schelling, T. Nobel Lecture: An Astonishing Sixty Years: The Legacy of Hiroshima. 2005. Available online: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2005/schelling-lecture.html (accessed on 11 February 2014).
3. *Handbook of Experimental Economics*; Kagel, J.H., Roth, A.E., Eds.; Princeton University Press: Princeton, NY, USA, 1995.
4. Camerer, C.F. *Behavioral Game Theory: Experiments in Strategic Interaction*; Princeton University Press: Princeton, NJ, USA, 2003.
5. Field, A.J. *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity*; University of Michigan Press: Ann Arbor, MI, USA, 2001.
6. Field, A.J. Group Selection and Behavioral Economics. In *Foundations and Extensions of Behavioral Economics: A Handbook*; Altman, M., Ed.; M.E. Sharpe: New York, NY, USA, 2006; pp. 165–182.
7. Field, A.J. Beyond foraging: Behavioral science and the future of institutional economics. *J. Inst. Econ.* **2007**, *3*, 265–291.
8. Field, A.J. Why multilevel selection matters. *J. Bioecon.* **2008**, *10*, 203–238.
9. Thaler, R.H.; Shefrin, H. An economic theory of self-control. *J. Polit. Econ.* **1981**, *89*, 392–406.
10. Field, A.J. Prosociality and the military. *J. Bioecon.* **2014**, *16*, forthcoming.
11. Zagare, F.C.; Kilgour, D.M. *Perfect Deterrence*; Cambridge University Press: Cambridge, UK, 2000.
12. Trachtenberg, M. *History and Strategy*; Princeton University Press: Princeton, NY, USA, 1991.
13. Kahn, H. *Thinking about the Unthinkable*; Horizon Books: New York, NY, USA, 1962.
14. Von Neumann, J.; Morgenstern, O. *Theory of Games and Economic Behavior*; Princeton University Press: Princeton, NY, USA, 1944.
15. Nash, J. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 48–49.
16. Poundstone, W. *Prisoner's Dilemma*; Doubleday: New York, NY, USA, 1992.
17. Rhodes, R. *Arsenals of Folly: The Making of the Nuclear Arms Race*; Alfred A. Knopf: New York, NY, USA, 2007.
18. Jervis, R. *The Illogic of American Nuclear Strategy*; Cornell University Press: Ithaca, NY, USA, 1984.
19. Royal Swedish Academy of Sciences. Robert Aumann's and Thomas Schelling's Contributions to Game Theory: Analyses of Conflict and Cooperation. 2005. Available online: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2005/advanced-economicsciences2005.pdf (accessed on 11 February 2014).
20. Schelling, T. *The Strategy of Conflict*; Harvard University Press: Cambridge, MA, USA, 1960.

21. Selten, R. Spieltheoretische behandlung eines oligopolmodells mit nachfragerträgeheit. *Z. Gesamte Staatswiss.* **1965**, *121*, 667–689.
22. Rapoport, A. Review of Schelling (1960). *J. Am. Stat. Assoc.* **1961**, *56*, 433–438.
23. Walt, S.M. Rigor or rigor mortis? Rational choice and security studies. *Int. Secur.* **1999**, *23*, 5–48.
24. Luce, D.; Raiffa, H. *Games and Decisions: Introduction and Critical Survey*; Wiley: New York, NY, USA, 1957.
25. Barkow, J.H.; Cosmides, L.; Tooby, J. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*; Oxford University Press: New York, NY, USA, 1992.
26. Schelling, T. Autobiography. Available online: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2005/schelling-bio.html (accessed on 11 February 2014).
27. Zagare, F.C. Reconciling rationality with deterrence: A reexamination of the logical foundations of deterrence theory. *J. Theor. Polit.* **2004**, *116*, 107–141.
28. Rosenberg, D.A. The origins of overkill: Nuclear weapons and American strategy, 1945–1960. *Int. Secur.* **1983**, *7*, 3–71.
29. Isaacs, J.; Downing, T. *Cold War: An Illustrated History, 1945–1991*; Little Brown: New York, NY, USA, 1998.
30. Kaplan, F.M. *The Wizards of Armageddon*; Simon and Schuster: New York, NY, USA, 1983.
31. Kaysen, C. Strategic Air Planning and Berlin. Source: National Archives, Record Group 218, Records of the Joint Chiefs of Staff, Records of Maxwell Taylor. 1961. Available online: <http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB56/BerlinC1.pdf> (accessed on 11 February 2014).
32. Schelling, T. *Choice and Consequence*; Harvard University Press, Cambridge, MA, USA, 1984.
33. United States National Security Council. 1963. A Study of the Management and Termination of War with the Soviet Union. National Archives, Record Group 59, Department of State Records, Records of Policy Planning Council, 1963–1964, Box 280, File “War Aims”. Available online: <http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB31/05-01.htm> (accessed on 11 February 2014).
34. Hobbes, T. *Leviathan, or the Matter, Forme, and Power of a Common-Wealth Ecclesiastical and Civill*; The Clarendon Press: Oxford, UK, 1651.
35. Schelling, T. Review of Rapoport (1964). *Am. Econ. Rev.* **1964**, *54*, 1082–1088.
36. Schelling, T. Comments at A Celebration of the Achievements of Nobel Laureate: Thomas C. Schelling, Kennedy School of Government, Harvard University. 5 October 2006. Available online: <http://forum.iop.harvard.edu/content/celebration-achievements-nobel-laureate-thomas-c-schelling> (accessed on 11 February 2014).
37. Rubinstein, A. A theorist’s view of experiments. *Eur. Econ. Rev.* **2001**, *45*, 615–628.
38. Raiffa, H. *The Art and Science of Negotiation*; Harvard University Press: Cambridge, MA, USA, 1982.
39. Walt, S.M. A model disagreement. *Int. Secur.* **1999**, *24*, 115–130.
40. Rapoport, A. *Strategy and Conscience*; Harper and Row: New York, NY, USA, 1964.
41. Putnam, H. The Place of Facts in a World of Values. In *The Nature of the Physical Universe*; Huff, D., Prewett, O., Eds.; John Wiley: New York, NY, USA, 1979.
42. Field, A.J. Microeconomics, norms, and rationality. *Econ. Dev. Cult. Chang.* **1984**, *32*, 683–711.
43. Schelling, T. *Arms and Influence*; Yale University Press: New Haven, CT, USA, 1966.

44. Lerner, M. *The Age of Overkill: A Preface to World Politics*; Simon and Schuster: New York, NY, USA, 1962.
45. Cheney, D.L.; Seyfarth, R.M. *Baboon Metaphysics: The Evolution of a Social Mind*; University of Chicago Press: Chicago, IL, USA, 2007.
46. Ross, L. The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process. In *Advances in Experimental Social Psychology 10*; Berkowitz, L., Ed.; Academic Press: New York, NY, USA, 1977; pp. 173–220.
47. Suskind, R. *The One Percent Doctrine: Deep inside America's Pursuit of Its Enemies Since 9/11*; Simon and Schuster: New York, NY, USA, 2006.
48. Gintis, H. A framework for the unification of the behavioral sciences. *Behav. Brain Sci.* **2007**, *30*, 1–16.
49. Lee, S.P. *Morality, Prudence, and Nuclear Weapons*; Cambridge University Press: Cambridge, UK, 1993.
50. Wilson, D.S.; Sober, E. Reintroducing group selection to the human behavioral sciences. *Behav. Brain Sci.* **1994**, *17*, 585–654.
51. Boehm, C. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*; Harvard University Press: Cambridge, MA, USA, 1999.
52. Wilson, D.S.; Wilson, E.O. Rethinking the theoretical foundation of sociobiology. *Q. Rev. Biol.* **2007**, *84*, 327–348.
53. Powell, R. *Nuclear Deterrence Theory: The Search for Credibility*; Cambridge University Press: Cambridge, UK, 1990.
54. Guth, W.; Schmittberger, R.; Schwarze, B. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **1982**, *3*, 367–388.

© 2014 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).