



Calhoun: The NPS Institutional Archive

Center on Contemporary Conflict

CCC PASCC Reports

2013-09

Rationality, Culture and Deterrence / September 2013

Knopf, Jeffrey W.

PASCC, Project on Advanced Systems and Concepts for Countering Weapons of Mass Destruction

<http://hdl.handle.net/10945/37070>



Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

**Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943**

<http://www.nps.edu/library>



Rationality, Culture and Deterrence

Dr. Jeffrey W. Knopf
Monterey Institute of International Studies

SEPTEMBER 2013 | REPORT NUMBER 2013-009



PASCC
PROJECT ON ADVANCED SYSTEMS AND CONCEPTS
FOR COUNTERING WEAPONS OF MASS DESTRUCTION



Rationality, Culture and Deterrence

Jeffrey W. Knopf, Ph.D.
Monterey Institute of International Studies

September 2013

This report is the product of collaboration between the Monterey Institute of International Studies, Naval Postgraduate School Center on Contemporary Conflict, and the Defense Threat Reduction Agency.

The views expressed herein are those of the author and do not necessarily reflect the official policy or position of the Naval Postgraduate School, the Defense Threat Reduction Agency, the Department of Defense, or the United States Government.

This report is approved for public release; distribution is unlimited.

**U.S. Naval Postgraduate School (NPS)
Center on Contemporary Conflict (CCC)
Project on Advanced Systems and Concepts for
Countering WMD (PASCC)**

Project Cost: \$82,782

PASCC Report Number 2013 009

The Naval Postgraduate School Center on Contemporary Conflict is the research wing of the Department of National Security Affairs (NSA) and specializes in the study of international relations, security policy, and regional studies. One of the CCC's programs is the Project on Advanced Systems and Concepts for Countering WMD (PASCC). PASCC operates as a program planning and implementation office, research center, and intellectual clearinghouse for the execution of analysis and future-oriented studies and dialogues for the Defense Threat Reduction Agency.

For further information, please contact:

The Center on Contemporary Conflict

Naval Postgraduate School
1411 Cunningham Road
Monterey, CA 93943

pascc@nps.edu

Table of Contents

EXECUTIVE SUMMARY	4
INTRODUCTION	5
WHY “MODELS” ARE RELEVANT TO DETERRENCE POLICY AND PLANNING	6
THE PARAMETERS OF TRADITIONAL DETERRENCE DEBATES	7
THE RATIONAL ACTOR MODEL: STRENGTHS AND WEAKNESSES	12
STRATEGIC CULTURE: STRENGTHS AND WEAKNESSES.....	16
A PLAGUE ON BOTH YOUR HOUSES? SHARED LIMITATIONS OF RATIONAL AND CULTURAL APPROACHES	20
SOCIAL CONSTRUCTIVISM: LOOKING FOR OVERLAP IN ACTORS’ VIEWS.....	25
DOMESTIC POLITICS: TRYING TO INFLUENCE INTERNAL DEBATES.....	27
PSYCHOLOGY: ANTICIPATING OTHERS’ BIASES AND RISK ORIENTATIONS.....	28
SUMMING UP THE ALTERNATIVES	34
CONCLUSIONS AND RECOMMENDATIONS	36

EXECUTIVE SUMMARY

Deterrence strategies involve trying to influence the decision-making of another actor. Because of this, efforts to determine whether to employ a strategy of deterrence or how to implement such a strategy require attempting to forecast what things will influence the other actor and how that influence will be exerted. There are several models or frameworks available that could assist with efforts to anticipate how another actor will be influenced. In practice, the most prominent public debates related to deterrence in the United States have tended to reflect two main approaches. People tend to assume either that the other side will behave like a rational actor or that it will be driven by a unique strategic culture. While both approaches have merit, extensive critiques have revealed that both also have significant limitations.

This research project reviews three other approaches that have been applied to studying deterrence: social constructivism, domestic politics, and psychology and neuroscience. None of these approaches, either alone or in combination, offers a perfect framework for predicting the outcomes of deterrence efforts. Each adds valuable insights, however, that are relevant to developing deterrence strategies. They draw attention to the degree of overlap in how the two sides understand key issues related to deterrence, to the range of internal debate on the other side, and to common biases and heuristics likely to affect the decisions of actors on the receiving end of deterrent threats. In particular, recent work has highlighted the dangers of pushing the other side into the domain of losses or tapping into underlying concerns about justice and fairness.

This study recommends that, in thinking about whether and how to deter other actors, analysts make use of all the different models available for anticipating how the other side will be influenced. This will not guarantee success in deterrence, but compared to relying on just a single framework for thinking about deterrence, use of multiple perspectives should reduce the chances of overlooking a critical flaw in deterrence planning. Making sure that policymakers and planners are familiar with a range of alternative approaches should therefore improve the U.S. ability to craft effective deterrent strategies.

INTRODUCTION

Far from being a relic of the Cold War, questions about deterrence are a recurring feature in U.S. policy debates. In the waning days of summer 2013, the U.S. government found itself wrestling with how to respond to a likely deterrence failure involving the use of chemical weapons in Syria. The Syria case raises important questions about deterrence. Was the Assad regime undeterrable, making it unwise for President Obama to have announced that CW use would cross a red line? Or did deterrence fail because the United States did not take the right steps to make its deterrent threat credible?

As with the Syria case, two big questions are typically involved in deterrence debates. First, can some other actor be deterred? Would an Iran with nuclear bombs, for example, be subject to deterrence? Second, if deterrence is possible, how is it best achieved? What will it take to deter the other actor from a particular course of action? This second question can be just as contentious as the first. Right up to the end of the Cold War, western analysts disagreed about what threats would most effectively deter the Soviet Union. Some believed that the ability to convey a risk that a significant percentage of the Soviet population and economy would be destroyed would be sufficient to induce caution in Soviet leaders. Others argued that the Soviet regime cared only about the survival of the Communist Party elite, meaning that deterrent threats would have to find ways to target those elites directly.

Public discussions of deterrence tend to revolve around two basic perspectives: treating the other side as a rational actor or instead as a product of a unique strategic culture. Participants in policy debates might not know or invoke the academic labels for these approaches, but their arguments tend to reflect the assumptions of one or the other. There are other approaches, however, to trying to understand human behavior and decision-making. The deterrence debates that take place among American elites have been truncated in nature. They are unnecessarily narrow, because they do not take full advantage of existing alternatives to rational actor and strategic culture models.

The following report has several inter-related objectives. First, it seeks to show that debates about deterrence among U.S. policymakers, defense intellectuals, and other elite commentators have been dominated by arguments that reflect rationalist and culturalist approaches. These are not the only approaches that inform deterrence debates, but they are

consistently more prominent than alternative ways of thinking about deterrence. Second, this report evaluates the strengths and weaknesses of both rational actor and strategic culture models. Both approaches have been the subject of strong critiques, and this report will summarize the limitations of both approaches. Third, because the two most common approaches to thinking about deterrence both have significant limitations, this report seeks to summarize several alternatives for analyzing deterrence strategy. It will review three alternatives: social constructivism, approaches based on domestic and bureaucratic politics, and behavioral approaches drawn from findings in psychology and neuroscience. None of these represents a new approach to deterrence in academic research. Instead, each of the three alternatives comes from a well-established body of work in International Relations (IR) and Security Studies. The goal here is to help policymakers, defense planners, and other relevant elites to become more familiar with these approaches. As with rational actor and strategic culture approaches, this report will summarize the strengths and weaknesses of the other three alternatives. It will show that none of the three alternatives, either singly or in combination with the others, offers a complete framework for analyzing deterrence. This report concludes that, in the absence of any fully reliable model, it will be useful to think in terms of multiple approaches to analyzing potential deterrence situations. The overarching objective for this report is to sensitize decision-makers and planners to the multiple ways available for thinking about deterrence and the value of incorporating insights from all of them.

WHY “MODELS” ARE RELEVANT TO DETERRENCE POLICY AND PLANNING

Although academic jargon is often opaque and the objectives of academic research quite different from what policymakers are concerned with, academic frameworks can still be relevant to the tasks that confront policymakers and planners. Practitioners often make decisions or develop plans based on implicit underlying assumptions about how the world works. These assumptions can be strikingly similar to those that inform academic theories or approaches. This is especially true in the case of deterrence. Because of what deterrence is about, general approaches for analyzing human behavior and decision-making have tremendous relevance.

Deterrence is about influence. A deterrence strategy seeks to influence the decision calculus of another actor. This makes it relevant to consider what influences people in general, and what specifically might influence the particular actor in question. From the perspective of somebody concerned about strategy, the ideal goal is the ability to predict the other side's behavior. Can the other side be deterred from a course of action? If so, what will be most effective in deterring it?

To predict whether another actor will be influenced, it helps to have some ideas about what makes people tick and how to influence them. Hence, an actor contemplating deterrence must make some assumptions about how others think. Without some idea about how to influence others, without some prediction of whether one's strategy will actually influence the other, it becomes hard to reach any conclusions about the merits of a potential deterrence strategy. Deterrence requires some attempt to predict what will influence the other actor, and for this reason people analyzing deterrent options will likely make some assumptions about the other side's decision calculus. The assumptions or ideas that guide thinking about deterrence usually derive from some general notion of what shapes the behavior and decisions of people, including collective actors such as states or terrorist organizations – and these general notions often resemble academic theories that deal with similar questions. For this reason, academic frameworks for understanding human behavior and decision-making could prove relevant to deterrence planning, and it will be helpful to be familiar with the full range of existing frameworks. The next section of this report summarizes the approaches that have historically had the greatest impact on thinking about deterrence.

THE PARAMETERS OF TRADITIONAL DETERRENCE DEBATES

Humans have practiced deterrence since time immemorial. It took the invention of nuclear weapons, however, to stimulate people to think deeply about deterrence. The major theories of deterrence are products of the nuclear age.

In the West, and especially the United States, deterrence theory initially developed primarily on the basis of a rational actor assumption. Three factors contributed to this. First, fortunately for humanity, the world lacks any clear evidence of how nuclear war breaks out. Ever since the Soviet Union joined the United States in the nuclear club, there has been no

use of nuclear weapons against another actor. In the absence of historical cases to study, it made sense to work with more abstract models to develop theories of nuclear deterrence. Second, in the early decades of the Cold War, Western analysts also found it difficult to obtain reliable information about decision-making in the highly secretive Soviet system. Without a clear understanding of how Soviet thinkers perceived nuclear weapons, relying on a generic rational actor assumption made it possible to develop some ideas about how nuclear deterrence might operate. Third, a rational actor framework also fit well with intellectual trends at the time. Tools such as game theory were first coming into vogue around the middle of the twentieth century, and it was natural that they would influence a subject involving “games” of strategy such as deterrence.

In academic Political Science, rational choice approaches have become associated with the use of formal models, such as extensive-form game theory models. In the most influential early thinking about nuclear strategy, however, analysts tended to employ the rational actor assumption more informally – they used it to think through basic requirements for deterrence. Examples include William Kaufmann’s work on the elements of credibility,¹ Albert Wohlstetter’s insights on the importance of deploying weapons in a survivable basing mode,² and perhaps above all the hugely influential writings of Thomas Schelling.³

By the 1960s, this approach had also begun to influence the U.S. strategic posture. Defense Secretary Robert McNamara’s adoption of “assured destruction” as the basis for determining the U.S. nuclear posture flowed from an assumption that a certain level of destruction would be adequate to deter any rational actor, including the Soviet Union.⁴ Familiar notions of crisis stability and strategic stability, and the mainstream theory of arms control, all took shape in this period on the basis of an informal rational actor assumption. Work in this

¹ William W. Kaufmann, “The Requirements of Deterrence,” in William W. Kaufmann, ed., *Military Policy and National Security* (Princeton, N.J.: Princeton Univ. Press, 1956).

² Albert Wohlstetter, Fred S. Hoffman, Robert J. Lutz and Henry S. Rowen, *Selection and the Use of Strategic Air Bases*, a report prepared for United States Air Force Project RAND, R-266 (Santa Monica, CA: RAND Corporation, April 1954).

³ Thomas C. Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960), and *Arms and Influence* (New Haven: Yale University Press, 1966).

⁴ Good histories of U.S. nuclear strategy include Fred Kaplan, *The Wizards of Armageddon* (New York: Simon and Schuster, 1983); Michael Mandelbaum, *The Nuclear Question: The United States and Nuclear Weapons, 1946-1976* (Cambridge: Cambridge University Press, 1979); and Lawrence Freedman, *The Evolution of Nuclear Strategy*, 3rd ed. (New York: Palgrave Macmillan, 2003).

tradition has become so widespread that it has been given its own label, rational deterrence theory (RDT).

By the 1970s, RDT had come under assault from a variety of critics. In the academic literature, many of the critiques focused on psychological constraints on rationality (this literature is discussed below). In the publicly most prominent debates about policy, however, the critique came from a somewhat different direction. Evidence about Soviet thinking on nuclear weapons had begun to emerge, and it troubled many observers. Soviet military writing about nuclear weapons did not reflect the same assumption common in civilian academic writing in the West that there was no meaningful way to use nuclear weapons to fight a war. The Soviet military buildup also seemed to defy predictions that the Soviet Union would accept the logic of parity and refrain from seeking a military advantage. Critics of mainstream U.S. thinking about nuclear strategy charged that, for reasons of history, geography, national identity, and the nature of the Communist system, Soviet leaders simply thought differently about nuclear war than did their Western counterparts. Historian Richard Pipes put forward the most widely read public statement of this position,⁵ but other analysts such as Colin Gray, Keith Payne, and Fritz Ermath also made the case that the U.S. and Soviet approaches to nuclear strategy diverged widely.⁶ Jack Snyder coined the term “strategic culture” to capture these differences,⁷ and that term stuck as the label for the primary public alternative to the rational actor model of deterrence (RAM).

This line of reasoning also had a significant influence on U.S. policy. It provided the basis for the Carter administration to adopt what it called the “countervailing strategy.” Echoing the academic critique, administration officials such as Defense Secretary Harold Brown argued that the United States could no longer take a “mirror image” approach and assume that threats sufficient to deter U.S. leaders would also deter Soviet leaders. Instead, U.S. nuclear targeting would hold at risk what the United States thought Soviet leaders valued

⁵ Richard Pipes, “Why the Soviet Union Thinks It Could Fight and Win a Nuclear War,” *Commentary* (July 1977).

⁶ Colin S. Gray and Keith Payne, “Victory Is Possible,” *Foreign Policy* (summer 1980); Fritz W. Ermath, “Contrasts in American and Soviet Strategic Thought,” *International Security* (fall 1978).

⁷ Jack L. Snyder, *The Soviet Strategic Culture: Implications for Limited Nuclear Operations*, RAND report R-2154-AF (Santa Monica: RAND, 1977).

most: the survival of Communist Party elites and their instruments for maintaining political control.⁸

The strategic culture approach has continued to influence deterrence strategy to this day. In the 2000s, the George W. Bush administration adopted a framework it called “tailored deterrence.”⁹ A version of this approach has been maintained by the Obama administration. As Jeffrey Lantis has astutely observed, the tailored deterrence strategy rests upon a strategic culture framework.¹⁰ U.S. deterrence planning documents do not base a tailored approach exclusively upon cultural factors and in practice seek to incorporate some of the other factors discussed below.¹¹ Yet the cultural approach is implicit in the basic assumptions of tailored deterrence. The tailored framework assumes that every country has a distinct and different national style. The goal of deterrence planning, therefore, becomes “to hold at risk what the other side values most.”

THE CONTINUING PROMINENCE OF THE TRADITIONAL POLES

When the Soviet Union disintegrated, the question of how to prevent major nuclear war became less pressing. Broader questions about deterrence did not go away however. New threats emerged, and with them familiar debates about whether they could be deterred and, if so, how. None of these debates revolves exclusively around rational versus cultural approaches, but in all of them there is a tendency for these to become the two most prominent poles in the debate.

The period before the U.S.-led invasion of Iraq in 2003 provides a good example. The intelligence failures concerning Iraqi WMD make it hard to remember the terms of the debate before the war. Prior to the invasion, the main axis of public debate was not whether

⁸ On adoption and nature of the countervailing strategy, see Scott D. Sagan, *Moving Targets: Nuclear Strategy and National Security* (Princeton: Princeton University Press, 1990); William E. Odom, “The Origins and Design of Presidential Decision-59: A Memoir,” in Henry D. Sokolski, ed., *Getting MAD: Nuclear Mutual Assured Destruction, Its Origins and Practice* (Carlisle, PA: U.S. Army War College, Strategic Studies Institute, 2004); Walter Slocombe, “The Countervailing Strategy,” *International Security* (spring 1981).

⁹ See M. Elaine Bunn, “Can Deterrence Be Tailored?” *Strategic Forum* (Jan. 2007) for a good short summary of this approach.

¹⁰ Jeffrey S. Lantis, “Strategic Culture and Tailored Deterrence,” *Contemporary Security Policy* (Dec. 2009).

¹¹ *Deterrence Operations Joint Operating Concept*, v. 2.0 (Offutt Air Force Base, Neb.: US Strategic Command, December 2006).

Saddam had WMD, but whether a Saddam armed with WMD could be contained and deterred. Supporters of launching a preventive attack on Iraq argued Saddam was a deterrence failure waiting to happen. Their arguments were not exclusively cultural; rather, they pointed to a mix of Saddam's personal tendencies to misperception, the lack of effective reality checks in the dictatorial Iraqi system, as well as a variety of factors reflecting Arab culture, recent history, and political dynamics.¹² Critics of this analysis argued largely from a rational actor perspective. They claimed that a shrewd political survivor like Saddam would understand that certain actions could prompt nuclear retaliation, and that he had nothing to gain from using WMD (or giving them to terrorists) that could rationally outweigh the costs of provoking a nuclear response.¹³

More recent debates about whether to attack Iran's nuclear facilities have been remarkably similar. Opponents of preventive attacks tend to assume that, should Iran cross the threshold to become a nuclear weapon state, it will soon find itself subject to the same logic of nuclear deterrence as all other nuclear-armed states, and it will quickly realize that it cannot use its nuclear weapons without inviting national destruction.¹⁴ Others argue that Iran is different. Because of elements in Persian culture and history, combined with the particular form of Shiite religious beliefs embraced by Iran's clerical leaders, the country's leaders might actually embrace martyrdom. According to this view, a nuclear Iran might become the first nation-state to act as a suicide bomber.¹⁵ Determining which view – if either – offers the better basis for predicting future Iranian behavior is hence a question for which the stakes could prove to be quite high.

The terms “rational actor model” and “strategic culture” represent academic labels. But the issues involved are not purely academic. Real-world debates about deterrence – whether it is possible, and how best to achieve it – often rest on arguments derived from these two frameworks. This makes it crucial to evaluate them. Is one consistently more reliable than the other? Is either an adequate guide to deterrence planning? Are these the only options, or are there other ways to think about deterrence that could be useful? The next two sections of

¹² Kenneth M. Pollack, *The Threatening Storm: The Case for Invading Iraq* (New York: Random House, 2002).

¹³ John J. Mearsheimer and Stephen M. Walt, “An Unnecessary War,” *Foreign Policy* (Jan./Feb. 2003).

¹⁴ For an extreme statement of this position, see Kenneth N. Waltz, “Why Iran Should Get the Bomb,” *Foreign Affairs* (July/Aug. 2012).

¹⁵ Adam Garfinkle, “Culture and Deterrence,” Foreign Policy Research Institute e-notes, Aug. 25, 2006.

this report analyze the strengths and weaknesses of rational actor and strategic culture approaches, respectively. Subsequent sections introduce several alternative approaches.

THE RATIONAL ACTOR MODEL: STRENGTHS AND WEAKNESSES

Multiple approaches for analyzing deterrence exist in part because it is hard to demonstrate that any one of them is unambiguously superior. Rather, each has distinct strengths and weaknesses. This section identifies some key advantages and disadvantages of a rational actor assumption. Because the literature on rational theories of deterrence is vast, this section makes no attempt to summarize the full range of research.¹⁶ This report also steers clear of the lively debate over whether rational actor approaches require the use of formal game-theoretic models. The question of concern here, rather, involves the policy-relevant strengths and weaknesses of assuming the other side is a more-or-less rational actor.

STRENGTHS

Conflict situations tend to push parties involved to make strategic calculations. As best as possible, they have to think ahead about how to advance their objectives and block the other side from imposing unacceptable losses on them. This provides the strongest argument for working with a rational actor framework: the assumption of rationality leads naturally to a focus on the element of strategic calculation in conflict situations. Whatever else is going on, some part of the reasoning process will probably involve judgments about whether a particular course of action is likely to succeed or fail and the gains and losses likely to follow from taking the action. These calculations can fruitfully be captured by thinking of the other side as a rational actor.

Even critics of rational deterrence theory can embrace this point. Colin Gray, for instance, is a leading proponent of the strategic culture approach. In the aftermath of 9/11, however, Gray rejected the widespread assumption that terrorism cannot be deterred. A common view emerged that sees suicide terrorists as irrational religious fanatics who want to kill themselves in order to reap promised heavenly rewards. Gray, and others, observed however that

¹⁶ For literature reviews, see Paul K. Huth, "Deterrence and International Conflict: Empirical Findings and Theoretical Debates," *Annual Review of Political Science*, Vol. 2 (1999); and Vesna Danilovic and Joe Clare, "Deterrence and Crisis Bargaining," in *International Studies Compendium*, ed. Robert E. Denmark (New York: Wiley-Blackwell, 2010).

terrorist groups such as al Qaeda have political goals and use terrorism strategically as a tool for trying to achieve their goals. This means they still think strategically, and as such they can in principle be deterred through efforts to deny them the ability to use terror to achieve their objectives.¹⁷

Related to this point, an assumption of rationality can be a useful safeguard against a tendency to stereotype an adversary. On the one hand, this can help guard against a risky assumption that the other side cannot be as smart and capable as one's own side. Treating the other as rational encourages a certain level of respect for potential adversaries that can be useful in getting one's own side not to let down its guard. On the other hand, as a corrective to stereotyping, the rationality assumption can also help analysts avoid giving up on deterrence prematurely should they have an inaccurate view of the other side as too fanatical or irrational to be subject to deterrence.

Advocates of rational actor approaches also highlight a number of advantages that should make this approach attractive to those wrestling with questions about deterrence. An assumption of rationality makes it possible to simplify a situation. One can strip away much of the detail of individual cases in order to focus on the underlying strategic logic of the situation. By permitting simplification, the rationality assumption makes it easier for an analyst to model a situation and think through the likely consequences of alternative courses of action.

Related to this, the rational actor approach also encourages logical rigor. In rational actor approaches, analysts usually have to make their starting assumptions clear and reason deductively to reach conclusions. This makes the underlying logic of the argument fairly easy to follow and makes it possible to check if conclusions have been reached in a logically consistent manner. Rigor does not guarantee good results. An analyst could still misconstrue the situation and work with a model that is a poor match to the actual situation. The need to reason logically, however, still provides a safeguard against sloppy thinking.

¹⁷ Colin S. Gray, *Maintaining Effective Deterrence*, Strategic Studies Institute monograph (Carlisle, PA: U.S. Army War College, Aug. 2003). See also Paul K. Davis and Brian Michael Jenkins, *Deterrence and Influence in Counterterrorism: A Component in the War on al Qaeda* (Santa Monica, CA: RAND, 2002), and Robert F. Trager and Dessislava P. Zagorcheva, "Deterring Terrorism: It Can Be Done," *International Security* (winter 2005/06).

A rational actor approach has one other important strength. It sometimes leads to valuable insights that are not obvious or even counterintuitive. In rational deterrence theory, one of the most profound insights has resulted from using an expected utility framework to analyze the choices of the other side. Expected utility analysis simply assumes that an actor has alternative options available and chooses the course of action that it expects to provide the highest net utility. Applied to deterrence, one can assume that the other side is choosing between two options: to attack, or not to attack. If it estimates expected costs and benefits for both choices, it chooses whichever option is relatively better. This has a significant implication. If the costs of not attacking are too high, that is if continuing with the status quo is perceived as highly negative, a rational actor might defy deterrent threats and launch an attack even if it views one's deterrent threat as credible.¹⁸ A defender can make sure the costs of an attack will outweigh the benefits for the other side, but if living with the continuation of current trends is seen as having even greater cost, it can still be rational for the challenger to attack.

This means that an effective deterrence strategy must take into account the costs and benefits for the other side of not challenging, that is of continuing to show restraint. In practice, U.S. planning guidance has recognized this insight. The *Deterrence Operations Joint Operating Concept* describes “encouraging adversary restraint” as a form of deterrence.¹⁹ It is not clear that the actions required to do this should properly be defined as deterrence, but they can contribute to shaping a context in which deterrence is more likely to be effective. Sometimes, it is useful to alleviate the other side's fears about what will happen if it continues to accept the status quo and/or offer it benefits for continued good behavior. This useful insight emerged from work in the rational deterrence framework. Even though there are significant limitations to the rational actor approach, this insight is worth holding on to, as it will continue to be valid even if one moves to incorporate elements of other approaches into the analysis of deterrence.

¹⁸ A possible historical example might be the 1973 Yom Kippur War. Sadat expected to lose the war with Israel. But starting a war and being defeated was still preferable to continuing a situation in which Egypt no longer possessed the Sinai.

¹⁹ *Deterrence Operations Joint Operating Concept*.

WEAKNESSES

Some weaknesses of rational deterrence theory are well known, because they are in many ways the starting point for alternative approaches. For those attracted to a strategic culture approach, rational approaches suffer from a tendency to assume all actors share the same values. Although rational choice approaches do not require that all actors have the same utility function, discussions of nuclear deterrence informed by a rational actor assumption seem in practice to assume that all states will be equally motivated to avoid societal destruction. This discounts the possibility that some dictators, particularly those motivated by an extremist ideology, will be willing to risk a lot of harm to their own populations in order to consolidate their hold on power and advance their ideological cause. As Colin Gray puts it, rational approaches typically work on the basis of a generic or universal rationality, thereby overlooking the possibility that different actors could make rational calculations but on the basis of different value rankings.²⁰

One attraction of a rational actor approach is that it permits analysts to think through logical possibilities in circumstances when they do not have much information about the other side. This creates a risk that analysts will fail to collect and analyze such information even if it is available. There will be a temptation to rely on abstract reasoning according to the rational actor model rather than go through the more labor-intensive task of empirical research. Because rational models appear to be powerful tools for prediction, they can create incentives that lead analysts or planners to forego the hard work that is sometimes necessary to learn more about the other side. In some cases, those concerned with deterrence might fail to obtain available information relevant to determining whether and how to proceed with deterrent efforts.

The greatest weakness of the rational actor approach, however, is the sheer inability of real human beings to live up to its requirements. This has been the central argument of psychological approaches, discussed more fully below. There are inherent constraints on the ability of both individuals and organizations to carry out all the steps in a procedural model of rationality. People lack the time and brain capacity to identify all the relevant options,

²⁰ Colin S. Gray, "The Reformation of Deterrence: Moving On," *Comparative Strategy* (December 2003).

collect and evaluate the information necessary to calculate each option's odds of success and failure, and accurately calculate the payoffs and risks associated with each option. In addition, there are well-documented sources of bias in human reasoning that make some degree of misperception and miscalculation likely.

Defenders of rational choice approaches argue that rationality is meant only as a simplifying assumption, not an accurate depiction of reality. According to them, the assumption that people behave "as if" they are rational utility maximizers can be justified if it leads to more accurate predictions than can be obtained through alternative assumptions.²¹ This might be valid if one is interested in average or equilibrium behavior across a large number of actors and cases. As a way to model a competitive market with thousands of producers and consumers, use of the rational actor assumption in Economics has been highly fruitful. But a typical policymaker or defense planner is not concerned only with the average outcome across thousands of cases. Someone with policy responsibilities will be most concerned with how a particular individual actor is likely to behave in a particular situation. And if significant deviations from rationally predicted behavior are possible or even likely, then working with a rational actor model becomes much more problematic.

Deterrence does work a lot of the time, and actors do sometimes act in ways consistent with the predictions of a rational actor model. The rational actor approach has useful advantages and should not be casually set aside. But given how hard it is for people to fulfill the requirements of the rational actor model, it could be risky to rely upon this model if viable alternatives are available. In U.S. foreign policy debates, a strategic culture approach has been the most visible alternative. Does this approach offer a superior basis for thinking about deterrence?

STRATEGIC CULTURE: STRENGTHS AND WEAKNESSES

The term "strategic culture" was coined in the 1970s, but the underlying ideas have been around for some time. Just as a rational actor model can be applied to topics other than deterrence, so too can a strategic culture model. For those who embrace this framework, much of the focus has been on military doctrine and ways of war – the basic premise being

²¹ Milton Friedman, "The Methodology of Positive Economics," in *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953).

that different states have distinct national styles in how they think about and use military force. From here, it is not much of a stretch to think that strategic culture might have implications for how an actor responds to deterrent threats as well.

The initial interest in strategic culture subsided as U.S.-Soviet relations improved and the Cold War came to an end. Interest in strategic culture revived in the 1990s, thanks in part to a lively debate set off by a critique of the earlier work written by Iain Johnston.²² Johnston drew on social constructivism and identified it as the theoretical foundation for what he labeled as a third generation in work on strategic culture. Johnston differentiated this third, constructivist generation from the original first generation, and argued the earlier work suffered from a host of conceptual and methodological problems. On behalf of the first generation, Colin Gray attempted to rebut Johnston's criticisms.²³ Ever since, a number of scholars have attempted to render strategic culture into a useful approach for research, in many cases seeking a workable middle ground between or synthesis of the first and third generations.²⁴

In practice, the policy community has tended to apply a notion of strategic culture that more closely resembles the work of the first generation. For this reason, the following assessment of strengths and weaknesses focuses on the strategic culture model as the first generation would apply it. It distinguishes this approach from social constructivism, which is also concerned with culture but thinks about it in a different way. Social constructivism is discussed separately below.

STRENGTHS

The strategic culture framework derives from an enduringly powerful insight. It reflects the old adage to "know thy enemy." This advice remains as valuable today as it was in Sun Tzu's

²² Alastair Iain Johnston, "Thinking about Strategic Culture," *International Security* (spring 1995).

²³ Colin S. Gray, "Strategic Culture as Context: The First Generation of Theory Strikes Back," *Review of International Studies* (Jan. 1999).

²⁴ See John Glenn, Darryl Howlett and Stuart Poore, eds., *Neorealism versus Strategic Culture* (Burlington, VT: Ashgate, 2004), and Jeannie L. Johnson, Kerry M. Kartchner, and Jeffrey A. Larsen, eds., *Strategic Culture and Weapons of Mass Destruction: Culturally Based Insights into Comparative National Security Policymaking* (Palgrave Macmillan, 2008). For good reviews of the literature, see Jeffrey S. Lantis, "Strategic Culture and National Security Policy," *International Studies Review* (December 2002), and Christopher P. Twomey, "Lacunae in the Study of Culture in International Security," *Contemporary Security Policy* (Aug. 2008).

time. In crafting a strategy to respond to a potential or actual adversary, an attempt to learn as much about the other side as possible should improve the odds of formulating an effective strategic response.

Seeking to do this also helps guard against potential weaknesses in the rational actor approach. Whereas a rational actor approach can encourage abstract, armchair reasoning, using a strategic culture framework encourages a decision-maker or planner to seek out information about the other side. This creates incentives to consult with historians, area specialists, and individuals with experience in the country or region in question. Including a broader range of advisors in the process should reduce the chances of misjudging how a particular course of action will play out.

The strategic culture approach is also intended to reduce the chances of engaging in mirror imaging. If the other side thinks in a significantly different way than elites in one's own country, the assumption of a shared rationality could lead strategy dangerously astray. This begs the question of whether the other side is really likely to behave differently, but it still seems prudent to ask this question explicitly rather than assume that the other actor will make the same choice that one would oneself.

WEAKNESSES

The strategic culture approach is intuitively appealing because it seems like common sense. Nevertheless, it has potentially significant liabilities. Most of these problems are not inherent in the concept of culture per se, but arise instead from strong human tendencies to apply the concept of culture in particular ways. These lead to several interrelated problems when people analyze deterrence through the lens of strategic culture.

People relying on a strategic culture approach tend to assume that cultures are basically national in nature. That is, they typically view the boundaries of culture as aligning with specific nations or ethnic identities. When they cite examples of strategic culture, the examples seem invariably to involve some particular national identity. Hence, applications of

the concept point to a distinctly Russian or Chinese or American or Arab way of thinking and acting.²⁵

In practice, however, the boundaries of culture are not so sharply drawn. Relevant cultures also exist both above and below the national level.²⁶ People can be part of larger transnational groupings. Those who are French or German can also see themselves as being European or even part of “the West.” In addition, states can contain more than one sub-culture, with these varying by generation, political party, or region. In the United States, for example, the South remains a distinct region, and the coasts differ in certain ways from the nation’s heartland.

If these different cultures encompass different strategic tendencies, this creates complications for an attempt to use the strategic culture framework. There are now competing strategic cultures, and an analyst who assumes the national culture predominates could be led astray. In Germany, for instance, political leaders might operate in an identifiably German manner, but they might also respond to security issues in a manner that better reflects a broader European strategic culture. To continue with the German example, a chancellor who was raised in the former East Germany might also have a rather different worldview than one who grew up in the West. This criticism accepts that culture matters. The point here is that analysts and planners can too easily move straight to focusing on a single, national-level culture and thereby overlook the other cultural strands that might also be at work.

A second, deeper problem follows from the first. Analysts who focus on a distinctly national culture can find themselves on a very slippery slope that quickly descends to crude stereotyping. A plausible observation about a relative difference between two populations can harden into a fixed belief that all members of some society are incapable of acting in a certain way or are hard-wired to behave in some other way. It is unfortunately all too common to hear people say about their adversaries that they “only respect force” or “don’t value human life as much as we do.” The problem here is not the desire to understand the

²⁵ This is quite pronounced in Gray, “Strategic Culture as Context.”

²⁶ Twomey emphasizes this in “Lacunae in the Study of Culture.”

other side's culture, but rather an unfortunate human tendency to understand it poorly and turn it into a misleading stereotype of the other.

Another problem, closely related to the first two, involves a tendency to treat culture too deterministically. Strategic culture approaches tend to assume that cultures took shape long ago, that they tend to remain fixed, and that they lead fairly directly to people's choices and behavior.²⁷ Social constructivists dispute these assumptions. They argue that culture can be quite malleable and fluid, and that cultures can sometimes change dramatically in a short period of time. This should induce caution against making deterministic predictions on the basis of a particular model of a nation's strategic culture.

Finally, and related to this last point, strategic culture approaches also tend to underestimate the impact of an actor's situation or strategic environment. Actors can sometimes find themselves compelled by circumstances to act in ways contrary to their normal cultural predispositions. This is, of course, the underlying assumption of rational deterrence theory, which assumes that a well-designed deterrent posture can induce an adversary to behave the way we want regardless of its underlying preferences. If a culturalist approach can rest upon observations about the weaknesses of the rational actor model, a rationalist approach can equally be put forward as a response to the weaknesses of a strategic culture approach. If each approach has strengths that address the weaknesses of the other, is it possible to fix their problems by combining them?

A PLAGUE ON BOTH YOUR HOUSES? SHARED LIMITATIONS OF RATIONAL AND CULTURAL APPROACHES

It is easy to see why both rational and cultural models are so attractive to analysts. They both have intuitively plausible foundations. And each can be seen as a way to address critical problems in the other. Unfortunately, simply trying to combine the two frameworks does not solve these problems. They are not logically complementary. The assumptions of the strategic culture framework imply that the predictions of the rational actor model will often be wrong, and vice versa. Simply stapling the two together is likely to produce confusion and contradictions, not clarity.

²⁷ These are core elements of Johnston's critique.

In addition, the debate between these approaches can be seen as a manifestation of what psychologists have labeled the fundamental attribution error. Attribution theory in psychology deals with how people think about the causes of an individual's behavior. Attribution theory identifies two basic options: situational and dispositional. The first suggests behavior results from a person's circumstances, the second that it reflects their innate nature. When people view another actor as an adversary, they tend to assume that good behavior by the other is forced on them by circumstances, while bad behavior is a product of the other side's character.²⁸

Traditional deterrence debates fit neatly into this framework. Rational deterrence theory emphasizes situational attributions. Through a proper application of deterrence, one actor can create circumstances that compel the other actor to behave. Rational theories lead to a degree of optimism about deterrence prospects because they suggest a state has the capacity to shape the situation so that it constrains the behavior of an otherwise bad actor. Cultural models reflect a dispositional focus. The other actor is dangerous, and deterrence is difficult, because the other is disposed to behave badly. Deterrence is not impossible, but it is not enough to threaten the other with some generically unfavorable circumstances. One must tailor the threat around circumstances that the other side will find especially unwelcome given its character.

Psychologists find that humans display systematic biases when they make attributions. When a potential adversary behaves well, we tend to overestimate our own role in creating a situation that led to good behavior. When the other acts badly, we tend to assume too quickly that this reflects their fundamental disposition and to discount the role of circumstances in driving their behavior, including the potential impact of our own behavior on the other actor. To the extent that rational actor and strategic culture approaches to deterrence align with situational and dispositional modes of attribution, respectively, psychology suggests that both are prone to error.

Beyond their close fit with well-established psychological dynamics, rational actor and strategic culture approaches have dominated thinking about deterrence for one more reason.

²⁸ For a good discussion of how this applies to foreign policy, see Alexander L. George, *Presidential Decisionmaking in Foreign Policy* (Boulder: Westview, 1980).

Both hold out the promise of making point predictions. Planners and decision-makers concerned about deterrence would ideally like to predict how a particular deterrence posture will influence other actors. Models that hold out the promise of making such predictions are hence inherently attractive. The deepest problem with both approaches, therefore, may be that such predictions are not possible, and the embrace of models that promise point predictions could prove to mislead decision-makers more than it helps them.

Rational actor models imply all actors are basically alike. Actors will be motivated by self-interest, and constrained by their circumstances, to make choices that maximize their utility. People might not make perfectly optimal choices, and there might be some minor variations across actors, but decisions are still likely to cluster narrowly around a uniquely best outcome. A rationalist approach to deterrence assumes that if the state practicing deterrence can credibly communicate an ability to make the outcome of challenging deterrence much worse than the outcome of not challenging, then we can predict that all self-interested actors are quite likely to make the same choice: to be deterred.

Strategic culture models imply that all actors are different. Each will make a unique choice reflecting its particular strategic culture. If an analyst understands that culture, however, the analyst can still make a point prediction. Each actor will behave in the way that reflects its nature. A culturalist approach assumes that if a deterrent threat is crafted to take account of the other side's nature, it will work, but otherwise it will fail. Across different actors, however, the appropriate deterrent posture could vary considerably.

By seeming to promise an ability to make clear predictions, both rational actor and strategic culture models have obvious allure for decision-makers and planners. Unfortunately, however, the bases on which both approaches promise point predictions are flawed, and relying exclusively on either approach could prove misleading. Three interrelated issues suggest why rationalist and culturalist approaches are both problematic. Figure 1 illustrates these points.

Why Traditional Approaches Were So Attractive – and Misleading

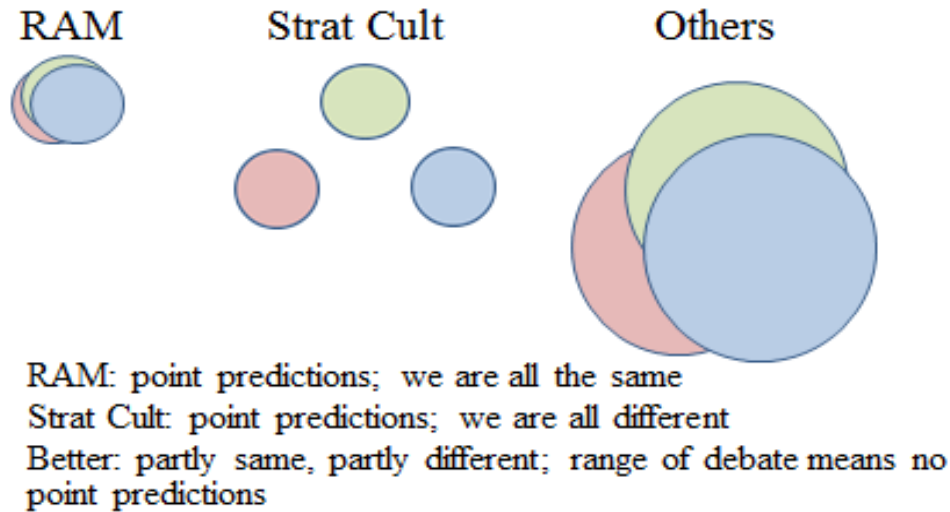


FIGURE 1: Rational, Cultural, and Other Approaches Compared

First, rational models assume that we are all the same, and cultural models that we are all different. It seems more accurate to assume, however, that we are partly the same and partly different. People across cultures still share many similarities. Some of these reflect common features of human nature. Others reflect the globalization process, in which people from different backgrounds are exposed to common aspects of world culture. Great works of art and literature often translate across cultural contexts because they deal with universal themes – love, loss, family, generational differences, triumphing over adversity – that we can all recognize and relate to. Yet we are not entirely the same, and art and literature can be used to make this point as well. Some things just do not translate. Those who emphasize strategic culture are correct to observe cultural differences. But they tend to focus only on the differences to the exclusion of the similarities. Rather than treat the existence of similarity or difference as a starting assumption, it might be more valuable to treat the degree of overlap across culturally different actors as an empirical question. Some actors might share more things in common, while others will have a greater divergence in their values and

worldviews. If actors are partly the same and partly different, it becomes difficult to predict either that they will all behave the same way or that they will all behave differently. Either outcome becomes possible.

Second, within any given actor, internal disagreement or debate remains possible and will often be observed. Even an individual leader can be internally conflicted and undecided. Most decision-makers must also consider the views of advisors, coalition partners, constituents, and possibly political opponents. Across such a larger group, internal disagreement and debate is even more likely. A particular culture that all the actors share might still contain competing norms or sub-cultures, and different actors will emphasize different ones. Uncertainty about the nature of the situation or what course of action is best, along with tradeoffs among competing values, also create scope for debate.²⁹ As a result of these factors, a range of different policy choices can appear reasonable. This makes point prediction much harder. If reasonable people can consider and debate a range of alternative choices, predicting their ultimate choice on the basis of either rational actor or strategic culture assumptions becomes quite hard. Many outcomes are possible, and other factors might shape the final choice.

Third, research in psychology has demonstrated that people can deviate in substantial and sometimes predictable ways from ideal rationality. Where the deviations are predictable, it might once again become possible to make point predictions, but the behaviors observed will differ from what both rationalist and culturalist approaches expect. The relevant biases and their effects are not always knowable in advance, however, so psychological approaches do not offer a reliable basis for making point predictions.

Collectively, these observations suggest that policymakers should anticipate a potentially wide range of possible behaviors by the actors they seek to deter, with the range of possible outcomes overlapping across actors but not entirely, and with final choices potentially influenced by psychological drivers that work against making fully rational decisions. These observations suggest that it will often be hard to predict with confidence the outcome of a particular deterrent effort. There is some modest good news, however. Each of the three observations about limits on the ability to make point predictions reflects an alternative

²⁹ George, *Presidential Decisionmaking*.

theoretical approach that can be applied to the question of deterrence. The next sections of this report briefly summarize the strengths and weaknesses of those three alternatives.

SOCIAL CONSTRUCTIVISM: LOOKING FOR OVERLAP IN ACTORS' VIEWS

Social constructivism is a general approach to understanding human social behavior. Since the 1990s, scholars have sought to apply this approach to the study of international relations. Like strategic culture, social constructivism emphasizes culture, but it conceives of culture in a different way. Constructivists do not link culture so closely to specific nations, but instead focus on commonalities across states. They also view culture as malleable and subject to change over the short term. Many constructivists emphasize forces in the international system that socialize states to become more alike in their ideas about appropriate state behavior.

In recent years, some scholars have started to explore the implications of constructivism for deterrence. The key observation is simple: deterrence is itself a social construct. To be able to play the deterrence “game,” actors have to share some common understandings. If there is no overlap in worldviews, constructing a deterrence relationship becomes harder.³⁰ This is similar to the core belief in the strategic culture approach that if actors have widely divergent value systems, it will be hard for one side to know how to deter the other. But constructivism puts the point in a subtly different way that has meaningful implications. Actors do not have to be culturally alike in all their values for deterrence to be feasible. Instead, they only need to share common understandings about key points relevant to deterrence. In particular, it helps if they share beliefs about the legitimacy of red lines and the appropriateness or proportionality of responses threatened if there are transgressions of those red lines.

Constructivists emphasize the importance of norms, and this extends to the realm of deterrence. When actors agree on norms, this makes them more likely to accept the

³⁰ Emanuel Adler, “Complex Deterrence in the Asymmetric-Warfare Era,” in T.V. Paul, Patrick Morgan, and James Wirtz, eds., *Complex Deterrence* (Chicago: University of Chicago Press, 2009); Amir Lupovici, “The Emerging Fourth Wave of Deterrence Theory—Toward a New Research Agenda,” *International Studies Quarterly* (September 2010).

legitimacy of deterrent threats linked to the violation of those norms. Terrorist groups are hard (but not impossible) to deter not so much because they have different value systems in general, but specifically because they view terrorism as a legitimate tactic and do not accept the efforts of other actors to define all terrorism as morally objectionable.

In relation to the diagram in figure 1, social constructivism draws attention to how much overlap there is in the understandings of different actors. The greater the degree of overlap, the better the chances for establishing a workable deterrence framework.

KEY POLICY IMPLICATIONS

If common understandings provide important underpinnings for deterrence, then efforts to establish common views can be an important policy tool. This includes measures that are not normally thought of as being related to deterrence. Anything that increases communication about issues related to a potential deterrence effort can help broaden the range of overlap in the views held by the two sides. This suggests that strategic dialogues can have unexpected benefits. They can help different actors come to a mutual understanding that certain actions constitute red lines and that certain types of responses are likely to follow. In addition to strategic dialogues, a number of other interactions could contribute to building up shared understandings. These include track II diplomacy, arms control talks, summit meetings, and even cultural exchanges. Actions that seem only distantly related to deterrence can contribute in unexpected ways to helping establish a context in which deterrence becomes more feasible.

WEAKNESSES

The key limitation of the constructivist approach is that the predicted effects of shared understandings are uncertain. As the overlap in actors' images of the world increases, this should improve the odds of successful deterrence, but it does not guarantee it. Actors can recognize the existence of a norm and still violate it. Syria's Bashar al Assad surely knows there is a widely supported international taboo on chemical weapons use, but he has still been willing to cross this particular red line as part of his attempt to hold on to power. Constructivism usefully draws attention to the ideational context in which deterrence efforts

unfold, but does not provide a silver bullet that can fix the limitations of rational actor and strategic culture models.

DOMESTIC POLITICS: TRYING TO INFLUENCE INTERNAL DEBATES

Analysts of international relations have long recognized that there can be significant disagreement inside a state or organization over what course of action to take. People can disagree for a wide range of reasons. In some cases, they have different parochial interests. In other cases, they might have different instincts concerning strategy, such as whether a hardline or conciliatory approach is likely to work best. The academic literature on deterrence has mostly not focused on how deterrence strategies might interact with domestic politics inside the target state. But a number of discussions of international bargaining more generally have taken this possibility into account.³¹

The basic rule of thumb these analysts have identified is simple and fairly intuitive. As a general rule, a state should seek to strengthen or at least not undermine the position of moderates in the other side's internal debate. And it should similarly seek to weaken or at least not reinforce the position of the other side's hardliners. What does this mean in practice? Suppose a relatively moderate faction has come to power in a state, but still faces significant competition from a more hardline faction. If the moderates reach out to seek a relaxation in tensions, a state should reciprocate that gesture and seek to reward the other side's moderates. Otherwise, it discredits them and gives ammunition to the argument of the other side's hardliners that one's own state is inflexibly hostile toward the other side. Conversely, if the other side's hardliners are in power and ratchet up the level of conflict, a state should stand firm or react with a modest increase in its own preparedness for conflict and not make unwarranted concessions. This can help the other side's moderates argue that their own state's aggressive behavior is responsible for provoking the conflict.

³¹ Graham T. Allison and Morton H. Halperin, "Bureaucratic Politics: A Paradigm and Some Policy Implications," *World Politics*, Vol. 24, Supplement: Theory and Policy in International Relations (spring 1972); Glenn Snyder and Paul Diesing, *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises* (Princeton: Princeton University Press, 1978); Robert D. Putnam, "Diplomacy and Domestic Politics: The Logic of Two-Level Games," *International Organization* (summer 1988); Jack Snyder, "International Leverage on Soviet Domestic Change," *World Politics* (Oct. 1989).

Needless to say, this is harder to do in practice than in theory. Efforts to influence internal debates inside another actor can backfire. If actors in another country believe a state is trying to interfere in its internal affairs, they can react with nationalistic pride against the effort and rally behind hardline policies. If a state seeks to craft a deterrence or other influence strategy to take account of internal debates in another actor, it must try to do so in a way that does not give the appearance of trying to change domestic political outcomes in the other side. The goal is to act in a way that gives credence to the arguments of moderates in the other side's internal debates without giving the appearance that the state wants to determine who actually wins political power in the other side.

Apart from the difficulty of implementing a strategy to influence internal debates in another actor, a domestic politics framework has one other potential weakness. It could lead an analyst to see greater domestic debate than actually exists. The appearance of debate could be misleading. It could be part of a ruse to lull a state into moderating its own stance when the other side actually has no intention of adopting a more moderate course. Or the debate could be real, but the moderate faction too weak to influence policy. If hardliners are fully in control and can quash or resist pressure from moderates for a change of course, then strategies predicated on affecting domestic debates inside the other actor are unlikely to work. Nevertheless, there are often internal disagreements inside actors, and strategic planning should not ignore such debates where they exist.

PSYCHOLOGY: ANTICIPATING OTHERS' BIASES AND RISK ORIENTATIONS

In the existing research on deterrence, approaches that draw on findings from the field of psychology – and more recently from neuroscience – are the best developed of the possible alternatives to rational actor and strategic culture approaches. There is no single unified or coherent psychological model. Instead, analysts have drawn on a range of different findings in psychological research to develop insights about how the human brain might affect deterrence outcomes. Much of this work has been concerned with highlighting flaws in the rational actor assumption.

THE INITIAL WAVE: SOURCES OF BIAS AND MISPERCEPTION

Starting in the late 1960s and early 1970s, critics began questioning the rationality assumption in mainstream deterrence theory.³² They turned to research in psychology to make the case that human decision-makers are not capable of fulfilling the requirements of the rational actor assumption.³³ The first wave of research emphasized certain well-documented biases in human reasoning that would be likely to produce misperceptions. These misperceptions, the literature suggested, could lead deterrence to fail even in situations in which deterrent threats should have been regarded as credible.

The research on psychology and deterrence ultimately sorted the various biases into two basic categories: cognitive (or unmotivated) and motivated.³⁴ The work of Robert Jervis³⁵ is probably the best known explication of cognitive biases, while Ned Lebow³⁶ made the case for the greater importance of motivated biases. Cognitive biases arise from the power of pre-existing beliefs and images. In short, people see what they expect to see. They try to make new information fit already existing beliefs. It therefore takes a great deal of discrepant information before people revise their beliefs. If leaders in state A view state B as irresolute, it will take a great deal of effort for state B to convince state A that it really will back up its deterrent threats.

Motivated biases basically involve wishful thinking. People see what they want to see. Here, decision-makers have some underlying need or desire that is driving their behavior, and they are motivated to interpret reality as being favorable to their needs. Take, for example, a situation in which the leader of a state felt he was losing support domestically and believed

³² Stephen Maxwell, "Rationality in Deterrence," *Adelphi Papers* no. 50 (1968); Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974).

³³ This has been a major theme in the work of Janice Gross Stein. See Janice Stein, "Deterrence and Reassurance," in Philip Tetlock et al. (eds.), *Behavior, Society and Nuclear War*, vol. 2 (New York: Oxford University Press, 1991), and "Rational Deterrence against 'Irrational' Adversaries? No Common Knowledge," in Paul, *Complex Deterrence*.

³⁴ Robert Jervis, Ned Lebow, and Janice Stein, *Psychology and Deterrence* (Baltimore: Johns Hopkins University Press, 1985).

³⁵ Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976).

³⁶ Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis* (Baltimore: Johns Hopkins University Press, 1981).

he needed to provoke and win a confrontation with a foreign adversary in order to rally the support to stay in power. In those circumstances, the leader would be motivated to believe the other side would back down, giving him the diplomatic victory he needs. Such a bias could lead a decision-maker to discount even highly credible deterrent messages from the other side. From the perspective of deterrence theory, this means a deterrent threat can fail even when the deterring side does everything right to make its posture credible.³⁷

For a policymaker or planner, the main drawback of this first wave of psychological literature is that it offered very little predictive ability. It is not easy to anticipate the particular biases or misperceptions that will affect the other side or to predict when these will be sufficiently severe to cause deterrence failure. Some research suggests that as the stakes increase people become more careful in their decision-making and misperception less likely.³⁸ Other research finds that high levels of stress degrade decision-making.³⁹ One can find historical examples of both results: Kennedy during the Cuban Missile Crisis proceeded very deliberately and cautiously, while Stalin seemingly had a nervous breakdown when Hitler's forces invaded the Soviet Union. It is hence hard to use the psychological approach as a basis for predicting when deterrent threats will succeed or fail. The main takeaway from the first wave was simply to sound a note of caution: deterrence can be fragile even when we think it should be robust and may be more prone to failure than rational deterrence theory would lead us to expect.

PROSPECT THEORY: PREDICTABLE DEVIATIONS FROM RATIONALITY

By the 1990s, IR scholars working on deterrence had discovered Kahneman and Tversky's work on what they came to label "prospect theory."⁴⁰ Prospect theory deals with how people make decisions in situations that involve risk. Based on a series of psychology experiments, Kahneman and Tversky demonstrated consistent patterns of deviation from the predictions of the rational actor model. Specifically, when given a choice between a relatively riskier

³⁷ Lebow, *Between Peace and War*.

³⁸ Daryl G. Press, *Calculating Credibility: How Leaders Assess Military Threats* (Ithaca: Cornell University Press, 2007).

³⁹ George, *Presidential Decisionmaking*.

⁴⁰ Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, XLVII (1979). For a good summary of the theory and its implications for IR, see Jack S. Levy, "The Implications of Framing and Loss Aversion for International Conflict," in Manus I. Midlarsky, ed., *Handbook of War Studies II* (Ann Arbor: University of Michigan Press, 2000).

option and a relatively safer option, people do not always make the choice that maximizes expected utility.

The most relevant finding from prospect theory is “loss aversion.” People are more motivated to avoid a loss than they are to make the equivalent gain. This, in turn, affects risk orientations. When humans find themselves in the domain of gains – that is, two choices that both promise a positive result – they become risk averse. Given a choice between a sure gain of a certain amount and a gamble that promises a bigger gain but at some risk of ending up with nothing, people choose the sure thing, even when it offers less expected utility than the alternative. In the domain of losses, however, this pattern reverses and people become risk acceptant. They will gamble on a possibly greater loss if that option also offers the possibility of escaping what would otherwise be a guaranteed loss of a lesser amount.

People identify gains and losses relative to a reference point, but identifying the reference point can be tricky. Often, it will be the status quo, but not always. The reference point can also be set by a desire to regain past glories or to achieve some future aspiration or to have what others have.

Prospect theory has potential implications for deterrence.⁴¹ First, it suggests that a state needs to be careful not to issue deterrent threats that move another actor into the domain of losses. Threats that are particularly harsh or broad or unexpected could convince another actor that it will end up with a bad outcome if it does not act, and this might lead it to take the risky gamble of challenging the deterrent posture in hopes of escaping from the shadow of the deterrent threat. Second, prospect theory suggests that the requirements for credibility may vary across domains. For a target state that is in the domain of gains, a deterrent threat does not need to be highly credible to be effective. Because actors are risk averse in the domain of gains, they will not want to take the risk of challenging even a moderately credible deterrent posture. If the other side is already in the domain of losses however, achieving a credible deterrent will be much harder because the other side will be inclined to accept the risk of challenging deterrence.

⁴¹ These have been explored by Jeffrey Berejikian, “A Cognitive Theory of Deterrence,” *Journal of Peace Research* 39, no. 2 (2002), and James W. Davis, Jr., *Threats and Promises: The Pursuit of International Influence* (Baltimore, MD: Johns Hopkins University Press, 2000).

In terms of crafting strategy, prospect theory suggests that efforts at framing the situation might be important. Gains and losses are defined relative to a reference point. If a state can help frame the interaction in a way that encourages the other side to see itself in the domain of gains, deterrence becomes an easier task. Successful framing, however, is likely to be quite a bit more difficult.

One advantage of prospect theory is that it holds out a promise of enabling predictions. If we can determine the other side's reference point and whether it sees itself in the domain of gains or losses, we can predict the likelihood that deterrence will work. In practice, however, prediction remains hard, for at least two reasons. First, ascertaining whether the other side sees itself in a domain of either gains or losses might be quite hard. This is not always true – countries that have recently lost territory or that face a serious internal threat to regime survival will almost certainly act like they are in the domain of losses. But in other cases this might not be as easy to predict. Second, and more problematically, the predictions of prospect theory are themselves only probabilistic. Some actors make choices contrary to what prospect theory predicts. The risk orientations it highlights are tendencies, not iron laws of human nature. Hence, an actor in the domain of gains could still surprise us and take a risky gamble. Even if it does not enable point predictions, however, prospect theory is valuable. It suggests some important rules of thumb about trying to avoid inadvertently pushing actors into the domain of losses.

NEUROSCIENCE AND BEHAVIORAL ECONOMICS: THE RETURN OF EMOTIONS

Traditional research in psychology has been supplemented by the emergence of new technologies that enable observation of the human brain as it operates. This work in the field of neuroscience has posed a new set of challenges to the assumption of rationality. Some of the previous research findings from psychology have been amalgamated with some of the findings from neuroscience in the school of thought known as “behavioral economics.” Economics has traditionally been the branch of the social sciences that most deeply embraced the rational actor assumption. Behavioral economics offers an alternative approach that draws on a wide range of findings about how the human brain really operates.

Much of the research that informs behavioral economics has been summarized for a general audience in Daniel Kahneman's recent book *Thinking Fast and Slow*.⁴²

With respect to international relations in general and deterrence in particular, three interrelated themes in the recent work stand out. First, most specialists now reject the idea that rationality and emotions should be treated as entirely separate and opposed forces. Considerable research suggests instead that the ability to make rational decisions depends to some extent on emotions. The emotional feelings people have about alternative outcomes feed into the process of assigning value – or utility – to those outcomes. When people have distinct likes or dislikes, these become part of the yardstick by which they evaluate which choices are better or worse.⁴³ When emotions and rational thinking come into conflict, however, emotions will often swamp rational thought and drive behavior.

Second, scholars now talk about the impact of “emotions” in the plural, rather than “emotion” in the singular.⁴⁴ This reflects the fact that people can experience different emotions. Several distinct emotional states are possible, such as happiness, sadness, or anger. Research shows that different emotional states have different effects. One of the most important distinctions involves the differences between anger and fear. Fear tends to induce caution, whereas anger makes people more likely to take risks and act without much concern for the consequences. This has potential implications for a strategy like deterrence. A deterrent threat that creates a measure of fear in the target, or that takes advantage of a preexisting fear, has a decent chance of working effectively. In contrast, deterrent threats that anger the target are more likely to backfire and provoke escalation.

Third, several lines of research suggest that people give significant weight to considerations involving fairness or justice.⁴⁵ When people believe they are treated unfairly, this creates particularly strong reactions. Experiments involving the well-known ultimatum game show that people consistently reject highly unequal distributions of benefit even when, in strictly

⁴² Daniel Kahneman, *Thinking Fast and Slow* (New York: Farrar, Straus and Giroux, 2011). See also Dan Ariely, *Predictably Irrational* (New York: Harper Collins, 2008).

⁴³ Jonathan Mercer, “Rationality and Psychology in International Politics,” *International Organization* (winter 2005).

⁴⁴ Janice Gross Stein, “The Psychology of Assurance: An Emotional Tale,” in Jeffrey W. Knopf, ed., *Security Assurances and Nuclear Nonproliferation* (Stanford, CA: Stanford University Press, 2012).

⁴⁵ For an exploration of the implications for nonproliferation, see Jeffrey W. Knopf, “Disarmament and Nonproliferation: Examining the Linkage Hypothesis,” *International Security* (winter 2012/13).

economic terms of expected utility, these choices actually leave them worse off.⁴⁶ This suggests that states should be careful to avoid deterrent postures that appear egregiously unfair or unjust given the prevailing beliefs of the other side.

This last point could have implications for extended deterrence. An extended deterrent that covers threats against an ally's long-standing homeland territory will raise few eyebrows. But efforts to protect an ally that has recently conquered new territory or that is engaged in domestic suppression of a particular ethnic or religious group might be harder to carry off. Other actors will likely regard such allied security guarantees as unjust and will be more likely to react with defiance than with compliance to such deterrent threats.

Recent work in neuroscience and behavioral economics has added depth and nuance to earlier research in psychology. It has not, however, brought about improved predictive power. Some research has suggested that experienced elites are not as susceptible to these psychological effects compared to the average person who takes part in psychology experiments.⁴⁷ This does not mean that veteran decision-makers will behave as perfectly rational actors. All human beings are subject to these psychological tendencies, just to varying degrees. This variation creates complications when it comes to using behavioral economics to make predictions however. Still, to the extent one can identify the particular bias or heuristic or emotional state that is likely to be uppermost in the minds of decision-makers on the other side, it might be possible to make some reasonable predictions about the likely effectiveness of different deterrent threats.

SUMMING UP THE ALTERNATIVES

Preceding sections have reviewed the potential of approaches based on social constructivism, domestic politics, and psychology and neuroscience to provide frameworks for deterrence analysis. None of the alternatives, taken either singly or in combination, offers a complete framework. They usefully supplement prevailing approaches that rely on rational actor or strategic culture models, but do not fully replace them.

⁴⁶ David A. Welch, *Justice and the Genesis of War* (Cambridge and New York: Cambridge University Press, 1993); Jonathan Mercer, "Emotional Beliefs," *International Organization* (winter 2010); Stein, "The Psychology of Assurance."

⁴⁷ Emilie M. Hafner-Burton, D. Alex Hughes, and David G. Victor, "The Cognitive Revolution and the Political Psychology of Elite Decision Making," *Perspectives on Politics* (June 2013).

The main limitation of the alternative approaches is that they are not highly predictive. To an extent, that was the point. Rationalist and culturalist assumptions promise greater predictive power than is possible. The other approaches help us understand why point predictions are generally not possible.

The value of learning about constructivism, domestic politics, and psychology is to sensitize decision-makers and planners to aspects of deterrence encounters they might otherwise overlook. Constructivism draws attention to the degree of overlap in two sides' worldviews that affect their ability to reach a mutual understanding on deterrence. Can they agree that certain actions constitute legitimate red lines and that certain responses to violations are appropriate? Thinking about internal politics highlights the potential for internal debates and encourages policymakers to consider the range of debate and the internal balance of power among competing factions on the other side. Is it possible to help tilt the balance in a favorable direction? In addition, consideration of domestic politics will help identify possible internal drivers of the other side's behavior that could lead to misperception in the form of motivated biases. Finally, research in psychology and neuroscience remind us of the many ways human beings can fail to act rationally and the apparently unavoidable impact of emotions upon decision-making.

The alternative approaches also have important policy implications. They reveal how actions that we might not traditionally associate with deterrence can help create the context for making deterrence stable and effective. Strategic dialogues, for example, can be a vehicle for building shared understandings that improve the prospects for deterrence. Efforts to create situations that place the other side in the domain of gains can reduce the likelihood of challenges to deterrence. Tacit alliances with the other side's moderates can reduce the influence of hardliners who might be pushing for actions that create a risk of conflict.

Much of U.S. deterrent planning has traditionally focused on targeting – placing “warheads on foreheads” as people in the military sometimes put it. This continues to be an important element of the deterrence planning process. But it is far from exhausting everything that needs to be taken into account when considering the use of a deterrence strategy. Broadening the range of perspectives that decision-makers consider will enable them to include a wider range of relevant factors in the process of developing deterrent postures.

CONCLUSIONS AND RECOMMENDATIONS

Deterrence strategies involve trying to influence the decision-making of another actor. Because of this, efforts to determine whether to employ a strategy of deterrence or how to implement such a strategy require attempting to forecast what things will influence the other actor and how that influence will be exerted. There are several models or frameworks available that could assist with efforts to anticipate how another actor will be influenced. In practice, the most prominent public debates related to deterrence in the United States have tended to reflect two main approaches. People tend to assume either that the other side will behave like a rational actor or that it will be driven by a unique strategic culture. While both approaches have merit, extensive critiques have revealed that both also have significant limitations.

This research project has reviewed three other approaches that have been applied to studying deterrence: social constructivism, domestic politics, and psychology and neuroscience. None of these approaches, either alone or in combination, offers a perfect framework for predicting the outcomes of deterrence efforts. Each adds valuable insights, however, that are relevant to developing deterrence strategies. They draw attention to the degree of overlap in how the two sides understand key issues related to deterrence, to the range of internal debate on the other side, and to common biases and heuristics likely to affect the decisions of actors on the receiving end of deterrent threats. In particular, recent work has highlighted the dangers of pushing the other side into the domain of losses or tapping into underlying concerns about justice and fairness.

This study recommends that, in thinking about whether and how to deter other actors, analysts make use of all the different models available for anticipating how the other side will be influenced. This will not guarantee success in deterrence, but compared to relying on just a single framework for thinking about deterrence, use of multiple “models” should reduce the chances of overlooking a critical flaw in deterrence planning. Making sure that policymakers and planners are familiar with a range of alternative approaches should therefore improve the U.S. ability to craft effective deterrent strategies. In addition, familiarity with alternatives to rational actor and strategic culture approaches also underscores the wisdom of an earlier evaluation of U.S. deterrence strategy. If the goal is to

influence the behavior of another actor, it is prudent to think about all the different ways available to achieve influence and not just narrowly in terms of deterrence.⁴⁸

⁴⁸ George and Smoke, *Deterrence in American Foreign Policy*.