

The Determinants of Uncertainty in International Relations

JEFFREY M. KAPLOW 
William & Mary

AND

ERIK GARTZKE
University of California, San Diego

Uncertainty about military power is widely considered an important determinant of international conflict, but research in international relations provides relatively little guidance about the origins of uncertainty. What factors influence the validity of actors' assessments of military capabilities? When would one expect uncertainty about military capabilities to be particularly high, or especially low? We examine a series of factors capable of explaining the sources of uncertainty in international relations, positing that the uncertainty of assessments is a function of both characteristics of the state being assessed and the overall strategic environment. We test our theory using new measures of uncertainty that we derive from analyzing publicly available estimates of national military capabilities. Our findings offer a novel set of implications both for research that relies on uncertainty as a key cause of conflict and for policymakers faced with evaluating foreign military assessments.

La incertidumbre que genera el poder militar se considera, en general, un determinante importante de los conflictos internacionales. Sin embargo, las investigaciones en materia de relaciones internacionales proporcionan relativamente poca información sobre los orígenes de dicha incertidumbre. ¿Qué factores influyen en la validez de las evaluaciones sobre las capacidades militares que realizan los actores? ¿En qué momento se puede esperar que la incertidumbre que provocan las capacidades militares sea particularmente alta o baja? Evaluamos una serie de factores capaces de explicar las causas de la incertidumbre en las relaciones internacionales, y planteamos que dicha incertidumbre de las evaluaciones se debe a las características del estado que está siendo evaluado y al entorno estratégico general. Para poner a prueba nuestra teoría, usamos nuevos indicadores de incertidumbre que obtuvimos a partir del análisis de las estimaciones de las capacidades militares nacionales, las cuales son de dominio público. Nuestras observaciones ofrecen un novedoso conjunto de implicaciones para la investigación que se basa en la incertidumbre como la causa principal de los conflictos y para las autoridades que se encargan de analizar las evaluaciones militares extranjeras.

L'incertitude sur la puissance militaire est largement considérée comme étant un déterminant important des conflits internationaux, mais les recherches en relations internationales ne fournissent que relativement peu d'indications sur les origines de cette incertitude. Quels sont les facteurs qui influencent la validité des estimations de capacités militaires effectuées par différents acteurs? À quel moment peut-on s'attendre à ce que l'incertitude sur les capacités militaires soit particulièrement forte ou faible? Nous avons examiné toute une série de facteurs pouvant expliquer les sources d'incertitude en relations internationales en émettant le postulat que l'incertitude des estimations était fonction à la fois des caractéristiques de l'État évalué et de l'environnement stratégique global. Nous avons mis notre théorie à l'épreuve en utilisant de nouvelles mesures d'incertitude tirées de l'analyse d'estimations publiquement disponibles des capacités militaires nationales. Nos résultats offrent un jeu inédit d'implications, que ce soit pour les recherches qui reposent sur l'incertitude comme cause clé des conflits, ou pour les décideurs politiques confrontés à l'évaluation des estimations militaires étrangères.

Uncertainty is widely considered to be a central driver of several important phenomena in international relations. Bargaining theories of war point to uncertainty about a state's

capabilities or resolve as a major cause of conflict onset (Blainey 1973; Morrow 1989; Fearon 1995). Realist theory debates whether bipolar or multipolar systems are more unstable, driven by uncertainty among the great powers about who is a threat to whom (Deutsch and Singer 1964; Waltz 1979). Uncertainty about the capabilities or intentions of adversaries may result from misperception, leading to security dilemmas between states and setting conditions for escalating arms races and an increased risk of war (Jervis 1976, 1978). Uncertainty over a state's economic policy may affect private investment and international currency flows (Feng 2001; Leblang 2003).

It is no surprise, then, that countries devote considerable attention to remediating uncertainty: collecting intelligence, sharing information, and refining assessments of the material capabilities, resolve, and intentions of other states and non-state actors. International assessments are undoubtedly important—they reflect the way states see the world and may play a role in determining a range of international policies. At the same time, these assessments enjoy

Jeffrey M. Kaplow is an Assistant Professor of Government at William & Mary.

Erik Gartzke is a Professor of Political Science and Director of the Center for Peace and Security Studies at the University of California, San Diego, where he has been a member of the research faculty since 2007. Previous permanent faculty positions include Columbia University in the City of New York (2000–2007) and the Pennsylvania State University (1997–2000). He has held temporary positions at Dartmouth University, the *Ecole des Affaires Internationales* (Sciences Po), the Naval Postgraduate School, UC Santa Barbara, and the University of Essex. Dr Gartzke received a PhD degree in political science from the University of Iowa in 1997.

Authors' note: Earlier versions of this paper were presented at the 2015 ISA Annual Conference, the 2015 Empirical Implications of Bargaining Theory Conference at Princeton University, and the 2016 Empirical Implications of Bargaining Theory Workshop at the University of Notre Dame. The authors thank Scott Abramson, Paul Poast, Kris Ramsey, Josh Rovner, Oliver Westerwinter, and our fellow workshop participants for many helpful comments. An online appendix with additional materials, including replication files, is available on the ISQ Dataverse at <https://dataverse.harvard.edu/dataverse/isq> or at www.jkaplow.net/research.

varying degrees of success in cutting through the pervasive uncertainty surrounding the plans, intentions, and potential actions of other states.

A body of research has sought to explain both the content and the impact of state assessments of other actors. This work has significantly advanced our understanding of how assessments are made and how they translate into foreign policies. Scholarship in this area, however, has often focused on prominent cases of misperception, deception, or other types of errors in assessment (Jervis 1976). While this attention to examples of errors is clearly important, an almost exclusive focus on mistakes at the expense of other tendencies may cause researchers to develop errant conclusions of their own. International assessments form the backdrop of foreign policymaking; they are being developed and acted upon all the time, and they appear to have an influence on state behavior well beyond the high-profile examples that draw the most scrutiny. A broader approach—examining drivers of international uncertainty in crises and out—may produce more durable insights about the ability of states to discern capabilities and the role that uncertain assessments play in international conflict and cooperation.

It may at first seem obvious that states should be uncertain about one another's military capabilities. After all, nations treat details of their own weapons systems, military organizations, and operational plans as closely guarded secrets. Details of a country's military power help to determine its fate in battle. Why would any country want another nation or non-state actor to correctly assess its military capabilities?

While there are important reasons to conceal military power, there are also strong reasons to want to share information about military might, if not indiscriminately, then at least enthusiastically. If other actors perceive a state as powerful, that state is more likely to have its way in international affairs without having to fight. Diplomats can wield military power to obtain concessions from adversaries (and even allies) in lieu of actual fighting. Indeed, the advantages of being perceived to be capable often cause nations to attempt to exaggerate their own military potential. Most political actors would like to appear to be powerful in most situations, even when they are not.

Uncertainty can thus creep into international assessments for at least two reasons. First, countries that have military advantages may seek to conceal them, particularly when advantages are critical for the battlefield, when actors expect to use their capabilities in the relative short term, and when revelation of capabilities is likely to allow an adversary to field more effective countermeasures. Second, countries that lack military advantages have strong incentives to pretend that military advantages exist, especially when engaged in or anticipating critical negotiations with other states, and when states are trying to deter or compel rather than fight.

This study seeks to evaluate the determinants of uncertainty in international relations. How and why does uncertainty about the capabilities of other states vary across both time and target? When are states likely to have more uncertain perceptions of the capabilities and intentions of others? Under what circumstances are states more or less confident in their own assessments? We theorize about several characteristics of states and the international environment that are likely to influence the tendency for states to form valid perceptions, and we test several plausible drivers of uncertainty using new measures developed for the project. Our findings indicate that the set of factors leading to uncertainty in international relations is quite complex—factors that have little effect on uncertainty in some contexts can lead to more or less uncertainty in others. Overall, assess-

ments tend to be more accurate as the target state increases in military capability, becomes more democratic, or joins an international alliance. Interestingly, these same factors seem to cause those estimating military capabilities to judge their assessment as more uncertain.

Our focus in this article is on the drivers of uncertainty about the capabilities of other states. There are, of course, other types of international uncertainty, including uncertainty about resolve or uncertainty about another actor's intentions. While these other forms of uncertainty are clearly important, we emphasize uncertainty about military capabilities for three reasons. First, assessments of military capabilities are in some ways much easier than assessments of intentions or resolve. Questions about capabilities usually have an answer: a state has some number of tanks or has spent some amount on new armaments, whether or not that answer is known to others. Even a state's highest leadership, however, may not know the answer to questions about resolve or intention prior to experiencing an international crisis.¹ Further, intentions and resolve are more fluid than capabilities, complicating assessment—a leader's intent can change with no warning, but building new aircraft takes time. Understanding uncertainty about intentions and resolve requires engaging the wide range of behavior states use to signal their intent. Questions about capabilities do not avoid these issues altogether, but they certainly reduce the complexity involved. It makes sense, then, in this initial attempt to understand the drivers of uncertainty in international affairs, to focus on the necessary and simpler task of addressing the causes of uncertainty over state capabilities.

Second, capabilities are fundamental to analyses of other forms of uncertainty. States make judgments about others' resolve or intentions based in part upon their demonstrated capabilities. States might telegraph their intentions, for example, by investing in costly military forces, expending resources to mobilize them, or risking their survival by placing them in harm's way. The manipulation of capabilities can generate a costly signal, allowing other states to infer an enemy's motives or resolve (Glaser 1997). Psychological approaches to international assessment also recognize the importance of a state's capabilities in influencing perceptions of national intent. While these signals may be filtered through, and affected by, the expectations of the observer, they still form the raw data that underly and are the basis for international assessments (Yarhi-Milo 2013, 2014).

Third, uncertainty about capabilities is likely to be closely related to other forms of uncertainty in practice. Not only do assessments of capabilities factor into broader inferences about states' intentions and resolve, but they also reflect the same barriers to accurate perception of an adversary. Closed polities with little transparency, for example, generate uncertainty over both capabilities and intentions, while the clear statements about a state's capabilities embodied in, say, international arms control agreements also contribute to more valid assessments of intentions and resolve. Our focus here on assessing uncertainty about military capabilities, then, is a necessary and informative step, taking us at least part of the way toward understanding the drivers of other forms of international uncertainty.

¹Treverton (2003) refers to a question that has an answer—whether the answer is known to the assessor or not—as a “puzzle,” while a question to which there is no answer is termed a “mystery.” Assessing a foreign state's military capabilities is usually a puzzle, while assessing a foreign state's intent is more often a mystery.

Sources of Uncertainty in International Relations

What factors are most likely to influence one state's uncertainty about another state's military capabilities? Existing theoretical frameworks tend not to address this question directly. Bargaining approaches and other rationalist models focus primarily on the mechanisms by which states can communicate their intentions, such as by sending costly signals (Fearon 1997; Meirowitz and Sartori 2008; Slantchev 2010). Other approaches—emphasizing psychological, behavioral, or bureaucratic factors—pay particular attention to the lenses through which leaders view signals, and how a given piece of information may be interpreted differently based on the experiences of the assessor or the context in which the information is transmitted (Jervis 1976; Holmes 2013; Yarhi-Milo 2013, 2014; Horowitz and Stam 2014; Mellers et al. 2015).

In this article, we examine uncertainty about capabilities as a monadic feature of a state; that is, we theorize that characteristics of a state or its strategic environment affects the level of uncertainty surrounding its capabilities. By design, we reserve for future research the somewhat more complicated question of dyadic uncertainty. This is not to suggest that all states are equally uncertain about a particular target—it seems clear that some states have more resources to devote to understanding the capabilities of other nations, and states may take a particular interest in the capabilities of their potential adversaries. We theorize, however, that the military capabilities of some targets are systematically more transparent than others, whoever the assessing state. Capturing the features of a state and its military forces that contribute to this uncertainty seems a necessary starting point in thinking about determinants of uncertainty generally.

Two categories of factors specific to a target state are likely to affect uncertainty about its military capabilities. Characteristics of the state being assessed, such as its economic and military strength, and the transparency of its government, either facilitate or complicate the task for other actors in evaluating its capabilities. An actor's strategic environment—its pattern of alliances, rivalries, and recent or ongoing international conflicts—may also contribute to the transparency or credibility of military capabilities. We address each of these categories in turn.

State Characteristics

The uncertainty surrounding a state's military capabilities is at least in part a function of the particular attributes of the state and its military apparatus. A key factor is the size and complexity of the state's military forces. It can be very difficult to assess small military forces, such as separatist groups or guerilla fighters. Small militaries are challenging to assess for purely pedestrian reasons—missing one tank in a two-tank army means a 50 percent error rate—but also because smaller armed forces sometimes lack the visible indicators that attend to larger militaries and that reduce uncertainty generally. Large militaries, however, can introduce additional uncertainty of their own. Complex force structures and command and control arrangements can be difficult for outsiders to evaluate, and it is easier to conceal sensitive programs in a multitude than in a cast of dozens.

International assessments of countries with greater economic strength may carry less uncertainty for several reasons. First, wealthier countries may simply be of greater strategic importance than poorer countries, leading intelligence services and nongovernmental analysts to devote more effort to assessing their military capabilities. Second,

states with strong economies may be better able to make high-profile international arms acquisitions—purchasing expensive military equipment from countries such as the United States or Russia. Such purchases are widely noticed and readily tracked, while trades for less expensive armaments—small arms purchases, for example—draw much less attention. Finally, poorer countries may build different kinds of militaries by necessity, opting for more soldiers over expensive alternatives such as mechanized divisions or air forces (Sechser and Saunders 2010). These capital-intensive projects may be more easily detected and assessed by outside observers.

More broadly, the types of military forces employed, and by extension the force posture or military doctrines adopted by states, are likely to affect uncertainty around overall capabilities. Some types of military assets—aircraft carriers, for example—are hard to miss, while others, such as mobile missile systems, are easy to overlook. Similarly, some capabilities are designed to be seen, even advertised, in the sense that their utility comes from an adversary's knowledge of that capability. The possession of nuclear weapons, for example, may affect uncertainty over even conventional military capabilities in two ways. First, there is significant overlap between nuclear weapons platforms and conventional forces (Gartzke, Kaplow, and Mehta 2014). For example, missile systems that carry nuclear warheads are similar to (and sometimes identical to) missile systems that carry conventional warheads, and submarines may carry both nuclear and conventional missiles. Nuclear weapon states that draw attention to these platforms to strengthen their nuclear deterrent may reduce uncertainty about their conventional capabilities at the same time. Second, nuclear weapon states may draw additional scrutiny to all forms of military capability simply because of their nuclear status.

In contrast, some capabilities are only useful if an adversary is not aware of them. Cyber-attacks, for example, often take advantage of vulnerabilities that can be easily corrected once states understand the vector of attack (Gartzke 2013). Revealing the capability to exploit weaknesses in cyberspace or elsewhere may thus have the effect of reducing or eliminating that capability.

The transparency of domestic political institutions may play an important role in revealing capabilities to outsiders. A public budget process both exposes details of military expenditures and provides clues about a country's long-range military planning. Competitive procurement procedures, which often call for the publication of detailed specifications of military hardware in the acquisition process, further reduce uncertainty about a state's present and future capabilities. Of course, closed political systems are not necessary for keeping secrets; democracies, too, work to conceal capabilities in many cases. However, the task of hiding significant military capabilities becomes costlier in the face of societal norms or legal requirements emphasizing transparency. We therefore would expect assessments of the military capabilities of democracies to be less ambiguous on average.

This discussion suggests a number of testable hypotheses about the effect of state characteristics on the uncertainty surrounding estimates of military capability, summarized in table 1.

Strategic Interactions

The strategic interactions of states—their patterns of conflict and cooperation—probably also affect uncertainty about military capabilities, although the direction of the

Table 1. State characteristics and uncertainty about military capabilities

Factor	Hypotheses
Size and complexity of military	Larger and more complex militaries will be assessed with more/less uncertainty
Economic strength	States with strong economies will be assessed with less uncertainty
Types of military capabilities	States with nuclear weapons will be assessed with less uncertainty
Domestic political institutions	Democracies will be assessed with less uncertainty

effect of conflict processes on uncertainty is not always clear. Alliances, for example, are intended to reduce uncertainty about the balance of power. They are formed, at least in part, to deter or compel by signaling the combined might of two or more nations that might otherwise be seen as autonomous entities. Some multilateral alliances, such as NATO, promote transparency about the military contributions of member states. Membership in international institutions may have a similar effect. Signatories to some international arms control treaties are required as a condition of the agreement to share the broad outlines of relevant military capabilities. These requirements are not always adhered to, of course, but these agreements almost certainly reveal more than they obscure.

At the same time, however, alliance ties might mitigate the need for states to reveal details about their own military forces. Waltz (1979), for example, discusses the substitutability of internal and external balancing (arming versus allying). Both demonstrated military capabilities and alliances can function as a strategic deterrent, and they may well be treated as interchangeable by state leaders. If an alliance partner already boasts transparent military forces capable of deterring aggression, then the existence of the alliance frees up other members to conceal their own capabilities. In this way, alliances—particularly those including powerful states—may actually increase uncertainty about the military capabilities of some members of the alliance.

Uncertainty about military capabilities is likely to be influenced by rivalries between states, and by patterns of conflict more generally. Bargaining theory sees international conflict in part as a revelatory mechanism, reducing uncertainty battle by battle (Slantchev 2003; Smith and Stam 2004). In this context, assessments of states that are more frequent participants in interstate conflict ought to exhibit less uncertainty. These states have had ample opportunity to display their capabilities, reducing uncertainty overall (Wagner 2000). States engaged in interstate rivalries may also face more scrutiny of their military capabilities, leading to greater transparency. Seen differently, however, a history of conflict constitutes evidence that there is more uncertainty about a state's military capabilities—this, after all, may be what led to conflict in the first place. This selection mechanism, then, suggests a possible association between greater uncertainty and conflict-prone states.

Psychological theories suggest that framing and expectations color our perception of international signals (Jervis 1976; Yarhi-Milo 2013, 2014). When states are seen as a threat, for example, ambiguous information is likely to be interpreted as more alarming than when states are seen as benign. If these theories are correct, then we would expect to observe systematic overestimation of military capabilities for states with a recent history of conflict. Note that psychological components of perception and “rational” factors can coincide, as both are likely to guide human attempts at estimating capabilities.

The strategic environment of states suggests several additional hypotheses about the uncertainty surrounding assessments of their military capabilities, summarized in table 2.

Measuring Uncertainty in International Relations

Our understanding of the determinants of uncertainty is complicated by our inability to measure the concept in a satisfactory way (Gartzke 1999; Slantchev 2004). In the absence of direct measures of uncertainty, empirically minded scholars have narrowed their focus to the strategic settings within which uncertainty may be isolated. States may be more uncertain about the resolve of an adversary, for example, when the adversary has just experienced a change in leadership (Wolford 2007; Rider 2013; Smith and Spaniel 2019). Mediators may have more knowledge of the private information of state parties when they are allied with or otherwise biased in favor of one of the parties (Kydd 2003; Boehmer et al. 2004). More uncertainty may be present at the system level as concentrations of power change (Singer, Bremer, and Stuckey 1972; Huth, Bennett, and Gelpi 1992; Bennett and Stam 2004), and at the dyadic level when states are at power parity (Reed 2003b). States may reduce this uncertainty as they gain experience with adversaries (Bennett and Stam 2004), make pacts with allies (Bearce, Flanagan, and Floros 2006), or increase economic interdependence generally (Reed 2003a). Outcomes of particular battles may also resolve uncertainty about the outcome of the war overall (Smith and Stam 2004), or contests may prove more durable in the presence of conflict processes that generate further uncertainty (Shirkey 2016).

While these studies have clarified the effects of uncertainty in specific contexts, their empirical strategies have limited utility for a more general understanding of international assessment. The proxies used in existing work tend to be context-specific, limited to a particular unit of analysis, or closely associated with other likely drivers of uncertainty. These indirect measures do not allow for broad empirical tests of the drivers of uncertainty.

In this section, we describe two new measures of uncertainty that can be applied more broadly to international assessment. We draw from estimates of the size of military forces detailed in the World Military Expenditures and Arms Transfers dataset (hereafter WMEAT), originally compiled by the US Arms Control and Disarmament Agency (ACDA 1966–1997) and more recently by the US State Department (1998–2019). While there are other datasets with similar characteristics, WMEAT is the most comprehensive in terms of geographic coverage, time span, and breadth of sourcing, drawing on a variety of both US government and non-governmental source materials.²

²For a comparison of military expenditure data from WMEAT and other sources, see Brzoska (1981) and Lebovic (1998, 1999). While WMEAT data were compiled by the US government, these are public assessments based largely on publicly available sources. These data may well be distinct from non-public (classified) sources used by the United States and others. We address this issue in more detail in the online appendix.

Table 2. Strategic interactions and uncertainty about military capabilities

<i>Factor</i>	<i>Hypotheses</i>
Alliances	States that are members of alliances will be assessed with more/less uncertainty
Patterns of rivalry and conflict	States with international rivals or a history of conflict will be assessed with more/less uncertainty

The WMEAT volumes are perhaps best known to scholars and analysts as a source of military expenditure data (Brzoska 1981; Singer 1988; Lebovic 1995, 1998, 1999). While military spending is a widely used proxy for military capability generally, in this study we instead focus on the WMEAT data series that estimates the number of armed forces personnel in each state. We examine the size of military forces rather than military expenditures for three reasons. First, force levels usually represent an important input into military spending estimates.³ It seems useful therefore to understand the uncertainty associated with these figures before addressing the more complex calculus of military spending that necessarily includes a number of additional assumptions.

Second, as a practical matter, examining WMEAT's military spending data across volumes is complicated by its treatment of military expenditures in constant dollar terms with different base years and (sometimes unpublished) conversion rates over time.⁴ While there are several accepted methods for dealing with these currency conversion challenges, each has trade-offs (Brzoska 1981; Lebovic 1995). Focusing instead on the size of military forces eliminates the risk that we will attribute to uncertainty over military capabilities what is really uncertainty over currency valuations.

Finally, and most importantly, the military force data in these volumes represent a more conservative testbed for understanding the determinants of uncertainty. One of our uncertainty measures relies on capturing revisions to estimates of military capability over time. The WMEAT data on the size of military forces are much more stable than military spending data—seeing significantly fewer revisions over time—and thus reduce the risk that we will attribute an effect to our variables of interest that in reality comes from mere noise in our measure.

Retrospective Uncertainty

When a single source makes overlapping estimates across multiple years, we can derive a *retrospective* measure of uncertainty. Most WMEAT volumes, for example, provide estimates for the previous ten years: the 1973 edition covers the years 1963–1973 and the 1983 edition covers the years 1973–1983. Thus, each new volume revisits and often revises the estimates given in previous years. The 1973 edition of WMEAT provides a roughly contemporaneous estimate of military expenditures in 1973, while the 1983 edition provides an estimate for 1973 that has been revised with the benefit of a decade of hindsight.⁵ The relative change between the contemporaneous estimate and an estimate for that same year revisited a number of years later is an effective proxy for uncertainty about the state's military capabilities in the original time period.

For example, the 1999 edition of the WMEAT data judged that Kazakhstan had 33,000 members of its armed forces in that year. In the volume released ten years later, however, the estimate of the size of Kazakhstan's military in 1999 had been revised to 55,000, an increase of 67 percent. The relative or percent change in estimates like this one over time can tell us something about the level of uncertainty surrounding the original judgment. The more dramatic the change over time in an estimate of military capabilities for a given state and year, the more uncertainty attended to the original estimate.

A body of prior scholarship leverages this feature of the WMEAT data as it applies to estimates of states' military spending. While this literature offers several complementary findings, we adopt a somewhat different empirical approach to draw conclusions about the effect of state characteristics and the strategic environment on the uncertainty of military assessments. Several authors compare WMEAT estimates with assessments of military expenditures from other sources, although they do not treat the magnitude of the difference in estimates between these sources as a measure of the underlying uncertainty around military assessments (Brzoska 1981; Lebovic 1998, 1999). These studies also do not address the role of state characteristics and the strategic environment in explaining intersource variation. Other work focuses on the correlation among successive revisions of a given WMEAT estimate and patterns in these changes over time, testing theories of organizational learning (Lebovic 1995). While our primary analyses address a different dependent variable, our descriptive statistics—discussed below—are consistent with the finding that WMEAT assessments converge on a final estimate over time (Lebovic 1995, 1998, 1999). Our analytical approach is most similar to Lebovic (1998), which treats the magnitude of revisions to military spending estimates as a measure of estimative error, but this work does not examine the explanatory effect of state characteristics and their strategic environment.

Retrospective measures capture a kind of uncertainty that may not be recognized by states at the time; this type of uncertainty is similar to that referenced in Donald Rumsfeld's famous quip about “unknown unknowns.” Officials may be confident in an assessment, only to find out later that the estimate did not reflect the true capabilities held by a state.

One assumption underlying retrospective uncertainty is that estimates move closer to the true value over time.⁶ That is, we assume that the WMEAT estimates of a given military capability made ten years after the fact are at least on average better than estimates made contemporaneously.⁷ This seems plausible in most cases. New information about

³See Noren (2003) for a discussion in the context of estimating Soviet military expenditures.

⁴On the challenges of intertemporal comparison of military expenditure data generally, see Brzoska (1995).

⁵Estimates for the most recent year in a WMEAT volume—the 1973 estimate in the 1973 edition of WMEAT, for example—usually do have the benefit of some reflection; the 1973 edition of WMEAT was actually released in 1974.

⁶Large swings in estimates probably reflect some level of uncertainty even if later estimates are not more accurate. The assumption of improving estimates over time, however, is useful in linking our retrospective measure to uncertainty about military capabilities at the time of the original assessment.

⁷The assumption that a revised estimate is better than the original is widespread in the literature. It is often implicit, as in the preferential use of revised WMEAT estimates in the Correlates of War's datasets (Singer 1988), but is sometimes made explicit—see, for example, Anderton's (1989) discussion of military expenditure data in the context of the arms racing literature.

military capabilities comes to light over time, allowing analysts to refine and improve their estimates. This may be because a state has employed a capability that had been hidden, or because the development of new technologies allows for more valid assessments.⁸ Sometimes countries themselves make reliable declarations about their own capabilities after the fact—prompted by alliance ties or international arms control agreements—allowing earlier assessments to be revised.

If later estimates of national capabilities in a given year are better than earlier ones, then we would expect to see a particular pattern in our data on military capabilities. An assessment of a given state's capabilities should change more in the years closer to the original estimate, with variation decreasing as the assessment converges on a final value. Our data do seem to behave in this way. WMEAT estimates of the size of armed forces in 1973, for example, changed by an average of about 4 percent between the 1973 WMEAT edition and the 1975 volume. Two years later, the estimates had changed by an average of about 2 additional percentage points, and two years further on by about 2 more percentage points. Between the 1979 and 1983 volumes, however, estimates changed by less than half a percentage point on average, seeming to converge over time. The convergence we find in WMEAT assessments of the size of military forces is consistent with the pattern identified in other work on WMEAT's military spending estimates (Lebovic 1995, 1998, 1999).

This empirical pattern could also reflect diminishing interest in or ability to update assessments over time. Lebovic (1998), for example, sees convergence on later estimates of military expenditures as evidence “that estimators mainly attend to data for more recent years (to keep volumes current); thus primarily reveal information pertaining to re-estimates for proximate years; seek to refine estimates rather than reconsider estimating assumptions; reluctantly tamper with older estimates for which the institutional memory is vague; and generate information through a search process that yields diminished marginal returns.” For our analysis to be useful, however, it need only be true that the latest revised estimate is, on average, a better assessment of the level of military forces held by a state than the first such estimate. We would expect recent assessments to be an improvement over older assessments even if the updating process is largely clustered soon after the original estimate—as it might be, for example, if interest in updating old assessments wanes over time. Assessments need not be continually adjusted over a ten-year time frame in order for our retrospective measure to be a useful proxy for uncertainty.

Retrospective measures reflect changing judgments of foreign capabilities over time, but they may also capture changes in organizational priorities or assessment methodology—both official changes in policies and procedures and unofficial changes in area of focus or preferred sources. For example, it may be that organizational changes in the Reagan administration led ACDA to devote less attention to the revision of prior estimates (Lebovic 1999); one could imagine a similar effect when ACDA was subsumed into the State Department at the end of the Clinton administration. Such changes, however, do not negate the usefulness of retrospective measures of uncertainty. In order to bias our results, these organizational changes or shifts in collection methods would have to be systematically associated

⁸In one well-known example, US intelligence assessments in the late 1950s overstated Soviet bomber capabilities. This “bomber gap” was ultimately corrected after U-2 reconnaissance flights showed far fewer Soviet strategic aircraft than US analysts had feared. See Garthoff (2004) and Prados (1982).

with the independent variables in our analysis. If changes in counting rules or procedures are unrelated to our explanatory variables, then they merely add to the noise in our data and make it less likely that we would see a significant result in our models. Put another way, if we are capturing organizational or methodological changes—and not just changing assessments themselves—in the WMEAT data, then our findings likely understate the true effect of these factors on uncertainty over military capabilities.

Large swings in military personnel estimates due to methodological changes may even reflect the underlying uncertainty we are attempting to measure. For example, one possible methodological change is the decision to use one set of source material rather than another. If the estimated military strength of a state changes significantly after shifting sources, that change may in part reflect an underlying uncertainty around the original estimate between different source materials.

To better understand whether organizational or methodological changes—formal or informal—are responsible for our results, we replicated our models of uncertainty with the addition of a variable representing the year of the WMEAT volume used to construct the retrospective measure (the more recent of the two volumes). Adding this variable did not change the results for our independent variables of interest, suggesting that our findings are not biased by organizational or methodological changes over time.

Coder-Assessed Uncertainty

In some datasets, the collectors of the data are able to indicate their confidence in a particular assessment, resulting in a *coder-assessed* measure of uncertainty. Several volumes of the WMEAT data, for example, mark some military expenditure figures with an “E,” indicating that this is only an estimate, or an “R,” indicating a more uncertain rough estimate. Data also are sometimes omitted altogether, where uncertainty is so high that no reasonable estimate could be made.

These coder assessments provide a useful measure of uncertainty, offering a rough proxy for the uncertainty that was likely to confront policymakers at the time. Put another way, coder-assessed uncertainty is a measure of “known unknowns.” This type of uncertainty focuses on the confidence that analysts had in a specific assessment, rather than the fundamental accuracy of underlying estimates.

Comparing Measures of Uncertainty

Differences between retrospective and coder-assessed measures of uncertainty represent an analytical opportunity—each measure focuses on a slightly different aspect of the underlying concept of uncertainty. The essential difference between these dimensions of uncertainty lies in whether analysts are aware of their own uncertainty. Coder-assessed uncertainty reflects such an awareness, while retrospective uncertainty does not. One or both of these measures are available in several substantive areas across multiple years for nearly all countries, making them well suited for general tests of the role and impact of uncertainty in international affairs.

We can visualize these two forms of uncertainty in a table analogous to the “confusion matrix” used in statistical classification to convey the accuracy of predicted outcomes. In table 3, the rows represent coder-assessed uncertainty. This can be thought of as a kind of prediction made by the analyst about their own uncertainty. The first row

Table 3. Coder-assessed and retrospective uncertainty

		Retrospective uncertainty	
		Low	High
Coder-assessed uncertainty	Not marked as uncertain	True confidence	False confidence
	Marked as uncertain	False doubt	True doubt

includes observations that are not marked by analysts as uncertain. In these cases, the analysts are relatively sure of their assessments. The second row includes observations that are marked by analysts as particularly uncertain. Cases fall within this row when analysts are relatively unsure of their assessments. The columns in the table represent retrospective uncertainty. To the extent that later estimates are closer to the true level of military capability than earlier estimates, retrospective uncertainty can be thought of as the outcome that coder-assessed uncertainty was trying to predict. The first column includes observations that saw little or no change between the contemporaneous estimate and the revised estimate—these are cases with low retrospective uncertainty. The second column includes observations with high retrospective uncertainty. These cases saw substantial changes between an early estimate and a later assessment for the same year.

The diagonal of the table from the upper left to lower right, shaded in gray, represents cases in which analysts' knowledge of their own uncertainty was borne out by the need to revise the initial assessment in later years. In the upper-left cell, the analysts had confidence in their assessment and estimates did not see major changes over time. In the lower-right cell, analysts lacked confidence in their assessments, and this doubt was justified—these assessments saw more significant changes with the benefit of hindsight.

The unshaded diagonal in the table, running from the lower left to the upper right, includes cases in which analysts misjudged their own level of uncertainty. In the lower-left cell, analysts were too pessimistic. They marked their own estimates as uncertain, but the assessments did not change much over time. This may be because the analysts actually had better insight than they thought at the time, or because the analysts were appropriately uncertain but no new information became available with which to revise the estimate.⁹ In the upper-right cell, analysts had false confidence in their judgments. While they did not believe their estimates were particularly uncertain, the assessment saw significant revisions with the benefit of hindsight.

If analysts are regularly aware of their own uncertainty, we would expect our observations to fall disproportionately along the shaded diagonal, populating the “true confidence” and “true doubt” cells. In fact, estimates marked by analysts as particularly uncertain were more than twice as likely to see greater than average revisions; for data in the bottom row of the chart above, observations are disproportionately in the lower-right cell (“true doubt”) rather than the lower-left cell (“false doubt”). The same relationship, however, does not hold for estimates without coder-assessed uncertainty. These observations were revised roughly in proportion with the overall sample. This descriptive analysis suggests that the presence of coder-assessed uncertainty—

an entry marked as particularly uncertain—provides some information about the likelihood that this estimate will be revised, while the lack of such a mark may tell us little about the assessment's validity.

Retrospective uncertainty over military capabilities is relatively high in our data. Drawing from 14 WMEAT volumes, we calculated the absolute percentage change between an early estimate of military capabilities and an estimate that benefits from hindsight. Estimates of the size of a state's armed forces changed by an average of about 11 percent between the early and the retrospective estimate. The original judgment was off by more than 50 percent in about 5 percent of cases in our data, and by more than 100 percent in nearly 2 percent of the estimates. Coder-assessed uncertainty, however, is relatively infrequent in the WMEAT data. For the years in our data in which WMEAT coders indicated levels of uncertainty, only about 7 percent of assessments of the size of a state's military were marked as particularly uncertain or omitted altogether.

Figure 1 illustrates both types of uncertainty derived from WMEAT data on the size of a state's armed forces.¹⁰ The vertical axis represents retrospective uncertainty—the relative change between an early estimate and an assessment that benefits from a number of years of hindsight. Retrospective uncertainty greater than zero denotes an overestimate—the original assessment was higher than the retrospective assessment—while a level of uncertainty less than zero indicates an underestimate. Each circle represents an estimate for a given country in a given year. Data points that are located further from zero (the dashed line in the figure) are thus more uncertain by the retrospective measure. Triangles denote states with military capabilities that were marked by WMEAT data collectors as particularly uncertain—as estimates or rough estimates.¹¹ The solid line in figure 1 plots the mean retrospective uncertainty over time. The mean level of uncertainty is usually less than zero—the average assessment of the size of a state's armed forces is an underestimate.

Often, the two measures of uncertainty agree. Triangles located far from zero on the vertical axis in figure 1 are cases in which analysts correctly viewed their estimates as uncertain; the many circles close to zero are cases of true confidence—that is, coders' confidence in their assessments was warranted. However, the figure also shows substantial variation between these two dimensions of uncertainty. Triangles close to zero on the vertical axis are cases in which coders flagged their assessments as uncertain, but these estimates were not revised substantially over time. Finally, there are a number of cases in which coders' confidence in their assessments was misplaced; they did not see the estimate as

¹⁰ Figure 1 includes assessments only for those years in which a retrospective measure of uncertainty was available: 1967–1978, 1980–1983, 1986–1993, 1998–1999, and 2004–2005.

¹¹ Coder assessments of uncertainty are not available for all years. Estimates of military size that were missing in a given year are omitted from figure 1. For clarity, we also omit data points for which the absolute level of retrospective uncertainty was greater than 50 percent, but these data are still reflected in the mean.

⁹ We may be more likely to see high coder-assessed uncertainty coupled with low retrospective uncertainty in assessments of highly secretive states—analysts are aware that their estimates are uncertain, but they are unable to resolve that uncertainty over time.

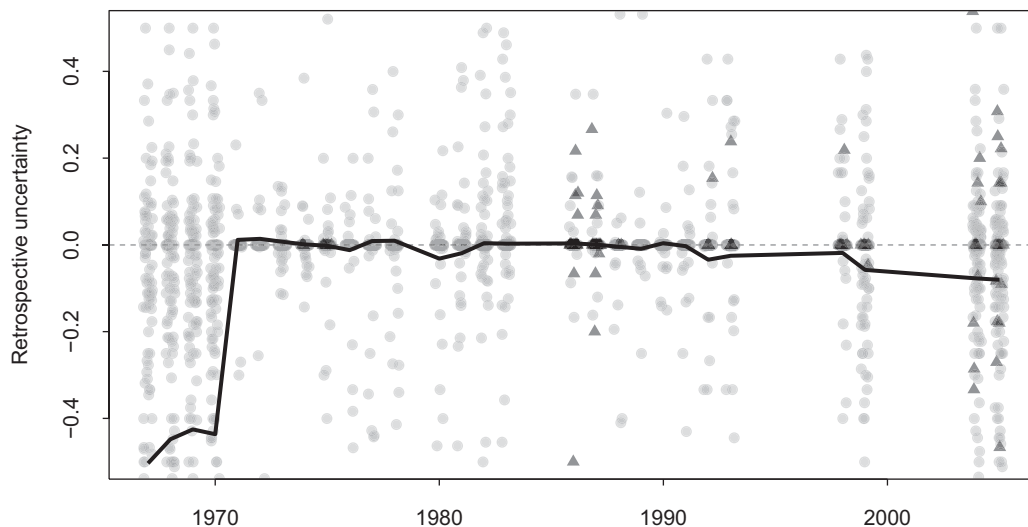


Figure 1. Retrospective and coder-assessed uncertainty over time.

uncertain, but it was revised substantially over time. These situations are represented in [figure 1](#) by circles far from zero on the vertical axis.

We have so far treated these measures as two independent dimensions of uncertainty. However, it is possible that coder-assessed uncertainty influences the future decision to revisit an estimate. Analysts may, for example, prioritize reanalysis of estimates that had been marked as particularly uncertain, believing these cases are most likely to yield substantial revisions. On the other hand, analysts might preferentially avoid revisiting these cases for several reasons: because such cases may be less likely to have new information available, because they represent more work for the analyst, or because analysts fear that revisions would signal greater confidence in an estimate than they actually have.

The potential interaction between these two forms of uncertainty threatens the validity of our findings on retrospective uncertainty. If estimates marked as particularly uncertain are less likely to be revisited, for example, our statistical models may miss an association between an explanatory variable and retrospective uncertainty that actually exists. Or, if coder-assessed uncertainty is associated with increased revisions of estimates in later years, our models may ascribe to retrospective uncertainty factors that are actually drivers of coder-assessed uncertainty. To better understand whether this is a problem for our analysis, we repeated our models of retrospective uncertainty with the addition of a variable representing the coder-assessed uncertainty measure. These robustness checks did not change our findings, suggesting that the relationship between these two forms of uncertainty does not confound our results.

Testing the Determinants of Uncertainty

To test the determinants of uncertainty about military capabilities, we construct a quantitative model, making use of time-series cross-sectional data with the country-year as the unit of analysis. The dependent variables in this analysis are measures of uncertainty about the number of personnel in a state's armed forces. We developed two measures of uncertainty based on estimates in fourteen volumes of the WMEAT dataset. To measure retrospective uncertainty, we used as a baseline the earliest available estimate for a given

state and year so long as that estimate was released no more than five years after the year being assessed. We then calculated the absolute percent change using the latest available estimate for a given year so long as that estimate was released at least eight years after the year being assessed.¹² This procedure resulted in twenty-eight years of data between 1967 and 2005.¹³

We employ three versions of the retrospective uncertainty measure. First, we calculate the absolute percent change for all observations. Factors that influence this variable can be thought of as affecting the accuracy of international assessments generally. This absolute measure of accuracy, however, might obscure some important conditional effects. It may be, as we have suggested, that some factors contribute to underestimation, while other factors cause collectors to overestimate capabilities. To assess this possibility, we subset our data by the direction of retrospective uncertainty. We thus analyze two additional dependent variables: absolute retrospective uncertainty among overestimates and absolute retrospective uncertainty among underestimates.¹⁴

To measure coder-assessed uncertainty, we created a dichotomous variable that takes on a value of 1 when the estimate of armed forces size in a country-year is marked by the WMEAT coders as particularly uncertain (i.e., marked with an "E" for estimated or an "R" for roughly estimated) or when an estimate is omitted from the WMEAT edition

¹² The years used for these cutoffs are the time between the actual preparation of the data and the assessed year, without regard to the sometimes arbitrary year mentioned in the title of the WMEAT volume. Our choice of cutoffs for early and retrospective estimates in the data was suggested by an analysis of the rate of convergence on a final estimate over time. While these cutoffs seem reasonable given constraints on the availability of data, our results do not depend on these parameters. More restrictive rules for including baseline assessments or final assessments yield similar results.

¹³ We lack a measure of retrospective uncertainty for 1979, 1984–1985, 1994–1997, and 2000–2003. That we could not analyze these particular years is purely a function of the availability of the WMEAT data in digital form. The editions of WMEAT that provided our data were not selected based on any judgment about the factors likely to lead to uncertainty in these years, and we have no reason to believe that these years are more likely than others to yield some association between uncertainty and our independent variables.

¹⁴ In all cases, we employ the absolute percent change from the contemporaneous estimate to the latest available estimate. By using absolute percent change rather than the absolute difference in terms of the number of military personnel, we avoid biasing the results for larger or smaller militaries.

altogether.¹⁵ Estimates are included in our data if they are the earliest available assessment released no more than five years after the assessed year.¹⁶ This procedure yielded thirty-two years of data between 1967 and 2010.¹⁷

The explanatory variables in our models attempt to capture potential drivers of international uncertainty. Because the size or complexity of military forces may affect the ability of other states to make accurate estimates, we include in our models the earliest WMEAT estimate of the number of personnel in the states' armed forces in that year. To capture the wealth and economic strength of the state, we include as an independent variable the state's real GDP per capita (Gleditsch 2002). Some types of military forces might be easier to assess than others—the fact that a state has nuclear weapons, for example, may simplify assessments of conventional military strength or focus attention on the details of military capabilities in that state. We therefore include in our analysis a dichotomous variable set to 1 if a state possesses nuclear weapons in a given year. We obtain data on nuclear status for each country-year from Gartzke and Kroenig (2009), updated to include North Korea as a nuclear weapon state since 2006 (Fitzpatrick 2011).

A state's regime type is another likely driver of the level of uncertainty about military capabilities. Democratic states may be more transparent regarding the make-up of their armed forces. Democracies are also more likely to subject military budgets to public debate, to have independent press reporting on military activities, and to engage with private enterprise in the procurement of military capabilities. We include in our models a measure of state regime type from the Polity IV project (Marshall, Jaggers, and Gurr 2010).

We account for the strategic environment faced by countries in several ways. First, we consider whether states are involved in an enduring rivalry, creating a dichotomous variable using data from Thompson and Dreyer (2011).¹⁸ Next, we account for the effects of alliance ties by using a dummy variable set to 1 if a state was a member of a formal defense pact in a given year (Gibler and Sarkees 2004). Finally, the task of assessing military capabilities may be affected by a state's history of conflict. States that have engaged in interstate disputes in recent years may well be seen as more threatening, with their capabilities drawing increased scrutiny from other states. Further, engaging in conflict may reveal a state's capabilities in more detail, making assessment easier. At the same time, however, states that anticipate future conflict have additional incentives to conceal military forces, and other states may view them with increased suspicion. We account for conflict history with a dichotomous variable set to 1 when a state participated in a militarized interstate dispute in the previous five years (Ghosh, Palmer, and Bremer 2004).

In time-series cross-sectional data of this kind, we are likely to see unobserved heterogeneity across units, leading to biased estimates in a fully pooled model. The usual alternative would be to include unit fixed effects in the model, but this limits our ability to interrogate variables with only cross-sectional variation or those that change slowly over time. Here, we adopt a within-between random

effects model, decomposing each explanatory variable into two measures—one capturing within-unit variation and the other between-unit variation (Bartels 2015; Bell and Jones 2015). We create the “between” variables by taking the unit-specific mean for each variable; the “within” measure then captures each observation's deviation from the unit-specific mean. This process leaves the within and between versions of the variables uncorrelated, allowing us to estimate the impact of our explanatory variables both cross-sectionally and within units. The within estimator in the within-between random effects model provides identical results to a traditional model that includes unit fixed effects (Bartels 2015; Bell and Jones 2015).

When the dependent variable is retrospective uncertainty, we model the impact of these various factors using linear regression. We model dynamics in these data by incorporating a lagged dependent variable as an explanatory factor (Beck and Katz 1996, 2011). This variable is the level of retrospective uncertainty associated with estimates of the size of a state's military in the previous year. A Lagrange multiplier test indicates that the inclusion of the lagged dependent variable effectively eliminates serial autocorrelation in these data (Beck and Katz 2011).

We model coder-assessed uncertainty using logistic regression. To address temporal autocorrelation, we include in this model a variable that counts the number of years since the WMEAT coders marked an estimate as particularly uncertain, along with squared and cubed versions of this variable (Carter and Signorino 2010).

Findings: Retrospective Uncertainty

Table 4 shows the results of our tests of the determinants of retrospective uncertainty. Model 1 examines retrospective uncertainty for the full data sample. The results for Model 1, however, may obscure interesting variation in the direction of the assessment error; we break out cases of over- and underestimation in Models 2 and 3, respectively. The table reports linear regression coefficients for these models, with robust standard errors clustered by country in parentheses. Positive and significant coefficients indicate that a given variable is associated with greater uncertainty; variables with negative and significant coefficients are associated with less uncertainty. Between-country results can be interpreted as the cross-sectional effect of that variable on uncertainty, while within-country estimates represent the effect of changes within a country over time.

Several state characteristics are associated with reduced uncertainty in international assessments. In Model 1, within-country effects for the size of a state's armed forces and for its real GDP per capita are negative and significant. The size of a state's armed forces is also associated with more accurate international assessments when the model is restricted to cases of underestimation. Increases in economic and military strength within countries over time are associated with a general reduction in retrospective uncertainty. The revelatory effect of large militaries and economies, or greater motivation to get those assessments right, seems to outweigh uncertainty engendered by complex military forces.

State acquisition of nuclear weapons, however, is not uniformly associated with reduced uncertainty. Within-state nuclear acquisition is linked to reduced uncertainty among underestimates, but to greater uncertainty in the subsample of overestimates. Put another way, joining the nuclear club appears to be associated with international assessments that are higher, on average. This finding is somewhat counterintuitive—nuclear weapons, after all, achieve

¹⁵ We only use omitted estimates for states that are themselves included in the WMEAT volume.

¹⁶ As before, more restrictive inclusion rules do not affect the results of the analysis.

¹⁷ We lack a measure of coder-assessed uncertainty for 1979, 1984–1985, 1994–1997, 2000–2003, and 2006.

¹⁸ Thompson and Dreyer's (2011) measure of rivalry does not rely on conflict density, and so captures a different dimension of interstate relations than our conflict history measure.

Table 4. Determinants of retrospective uncertainty

		(1) Retrospective uncertainty, all estimates		(2) Retrospective uncertainty, overestimates		(3) Retrospective uncertainty, underestimates	
		Between- country	Within- country	Between- country	Within- country	Between- country	Within- country
State characteristics	Size of armed forces	0.030 (0.039)	-0.338 (0.119)	-0.053 (0.146)	-0.231 (0.425)	-0.250 (0.268)	-2.331 (0.609)
	Real GDP per capita	-0.006 (0.010)	-0.013 (0.003)	-0.187 (0.105)	-0.122 (0.190)	-0.047 (0.047)	-0.002 (0.004)
	Nuclear weapons	0.001 (0.004)	0.005 (0.005)	-0.020 (0.036)	0.210 (0.031)	-0.012 (0.030)	-0.056 (0.026)
	Regime type	-0.021 (0.029)	0.036 (0.050)	-0.031 (0.153)	-0.414 (0.277)	-0.001 (0.002)	-0.013 (0.005)
	Strategic environment	Rivalry	-0.003 (0.004)	-0.009 (0.012)	-0.010 (0.028)	-0.005 (0.023)	-0.067 (0.037)
Alliance		-0.007 (0.004)	0.018 (0.012)	-0.075 (0.027)	0.053 (0.041)	-0.065 (0.036)	-0.064 (0.128)
Conflict history		-0.008 (0.006)	0.017 (0.012)	-0.072 (0.037)	0.028 (0.025)	0.036 (0.042)	0.072 (0.045)
Constant			0.023 (0.006)		0.242 (0.044)		0.147 (0.043)
	N		3,597		379		560

Notes: Linear regression coefficients with robust standard errors, clustered on country, reported in parentheses. Within-country coefficients are equivalent to a model including country fixed effects. Bold values are statistically significant ($p < .05$). A lagged dependent variable is included in all models but not reported.

deterrence in part through transparency. However, it may be that nuclear acquisition leads to greater scrutiny from analysts and thus systematically higher estimates, or that a greater sense of threat or perceived strength brought about by nuclear weapons colors analytic judgments in other military domains.

Shifts toward democracy are associated with a reduction in retrospective uncertainty when the sample is limited to underestimates. This finding, however, deserves a bit more scrutiny, because analysts may simply be less motivated to update estimates of democracies than they are of non-democracies. If this dynamic is present in the data-generating process, then we may mistakenly ascribe to uncertainty what is really a function of selective analyst interest in updating their assessments.

A closer look at the data suggests that this alternative mechanism is not confounding our results. If analysts are devoting less effort to updating democratic states versus non-democratic states, we might expect that democracies would be less likely to see any update at all. Uninterested or overworked analysts would be tempted to simply repeat the same estimate for a given year in subsequent WMEAT volumes. However, the data show that force estimates of democracies were updated at a slightly higher rate than non-democracies. About 31 percent of democratic observations saw revisions, compared to about 26 percent of non-democracies.

A more systematic way of investigating this alternative mechanism is to control for regime type at the time of re-assessment. If the alternative mechanism is correct, we would expect states that are more democratic at the time of re-assessment to exhibit smaller revisions to their estimates. To check this, we replicated Model 3 with the addition of this variable and found that a measure of regime type at the time of re-assessment does not reach statistical significance. In addition, our measure of regime type at the time of the initial

assessment yields the same result in this robustness check. This finding increases our confidence that we are capturing the effect of uncertainty around the initial estimate of a state's capabilities, rather than the selective interest of analysts in updating their assessments.

Among variables representing the strategic environment, the between-country estimator for interstate alliances is negative and statistically significant in Model 2. Compared to states without alliances, international assessments of allied states see less overestimation. Here again, the model suggests that the revelatory mechanism of alliance ties wins out over the use of alliances to obscure state military capabilities.¹⁹ It is difficult to unravel the impact of uncertainty on alliances and military conflict, however. A state may be more likely to initiate a dispute with an adversary if it underestimates the military power of the target or alliance coalition. Conversely, a state should be less likely to attack a country with strong allies. Alliances may be formed in anticipation of threats and threats may reveal themselves only in cases where alliances are weak. Adversaries are likely to evaluate and be affected by all of these forces simultaneously.

The capabilities of states are more underestimated when those states become engaged in international rivalries, a feature of uncertainty about military capabilities that is theoretically associated with increased conflict. States that are more prone to conflict have, in effect, already demonstrated that uncertainty attends their capabilities, since this may have led to increased conflict in the first place. Unlike states that have engaged in military conflict in recent years—the variable representing involvement in disputes is not significant in any model of retrospective uncertainty—states with rival-

¹⁹This finding does not solely capture the effect of NATO membership. Results for variables of interest do not change when a variable representing NATO membership specifically, rather than defense pacts generally, is included.

Table 5. Determinants of coder-assessed uncertainty

		(4) Coder-assessed uncertainty	
		Between-country	Within-country
State characteristics	Size of armed forces	22.170 (7.275)	-6.183 (3.704)
	Real GDP per capita	-1.100 (0.624)	-0.519 (0.680)
	Nuclear weapons	-6.162 (2.430)	2.551 (0.897)
	Regime type	-0.013 (0.020)	0.099 (0.030)
Strategic environment	Rivalry	0.016 (0.242)	0.328 (0.550)
	Alliance	0.214 (0.237)	2.206 (0.408)
	Conflict history	0.472 (0.358)	0.714 (0.309)
	Constant		3.960 (0.883)
	<i>N</i>		4,448

Notes: Logistic regression coefficients with robust standard errors, clustered on country, reported in parentheses. Within-country coefficients are equivalent to a model including country fixed effects. Bold values are statistically significant ($p < .05$). A cubic polynomial of years since previous coder-assessed uncertainty is included but not reported.

ries may not have had to reveal private information through the conduct of conflict.

Findings: Coder-Assessed Uncertainty

Table 5 presents results for our model of coder-assessed uncertainty; the table displays logistic regression coefficients with robust standard errors clustered by country. Here again, positive and significant coefficients reflect greater uncertainty, while negative results are associated with less uncertainty. Our findings represent an interesting contrast both between retrospective and coder-assessed measures of uncertainty, and between the two different types of random effects estimators in our models.

The between-country estimator for the size of the state's armed forces is positive and significant; compared to states with smaller militaries, larger armed forces are more likely to be marked as uncertain. Movement toward democracy within a country, the formation of international alliances, and recent involvement in militarized disputes are also associated with increased coder-assessed uncertainty. These results largely suggest the opposite relationship from our earlier models of retrospective uncertainty: When it comes to the retrospective accuracy of international assessments, the size of armed forces, the formation of alliances, and movement toward democracy are associated with less uncertainty. But of course, our retrospective and coder-assessed measures seek to capture different aspects of what we mean by uncertainty. Analysts are not always good arbiters of their own success, and the same factors that lead to less retrospective uncertainty—increased international scrutiny due to the strategic significance of a target state, for example—may cause international observers to have lower confidence in their judgments of a state's capabilities.

In Model 4, the between-country estimator for nuclear weapons capabilities is negative and statistically significant, while the within-country estimator is positive and significant. Compared to non-nuclear countries, coders are

less likely to express uncertainty about their judgments of nuclear weapon states. This finding is consistent with the role of nuclear forces as a military capability tied to strategic deterrence, whose effectiveness necessitates considerable transparency. At the same time, when a state crosses the nuclear threshold, analysts become more likely to mark estimates of its military capabilities as particularly uncertain.²⁰ This may be, in part, an artifact of the particular nuclear weapon states that are present in our data. The P-5 nuclear weapon states exhibit no within-country variation in nuclear weapons status in our dataset because they acquired a nuclear weapons capability before 1967. However, less established nuclear states—such as Pakistan and North Korea—have observations both before and after nuclear acquisition. It is also possible that nuclear acquisition creates a more complex environment for military assessments, leading to greater uncertainty over estimates of conventional military capabilities.

Substantive Effects

We find that a number of the determinants of uncertainty over state military capabilities are both statistically and substantively significant. A substantive interpretation of the linear regression results from Models 1–3 is straightforward. Among underestimates, for example, the creation of an international rivalry is associated with about a 15 percentage point increase in retrospective uncertainty, while among overestimates a state in an alliance has 7.5 percentage points lower uncertainty on average than states lacking an alliance. Figure 2 illustrates the substantive effects of our three continuous explanatory variables—size of armed forces, real GDP per capita, and regime type—on retrospective uncertainty in each of the three samples (all estimates, overestimates, and underestimates). For simplicity, only

²⁰This finding illustrates the utility of a within-between modeling approach; only the within result would have been observable with a fixed effects estimator.

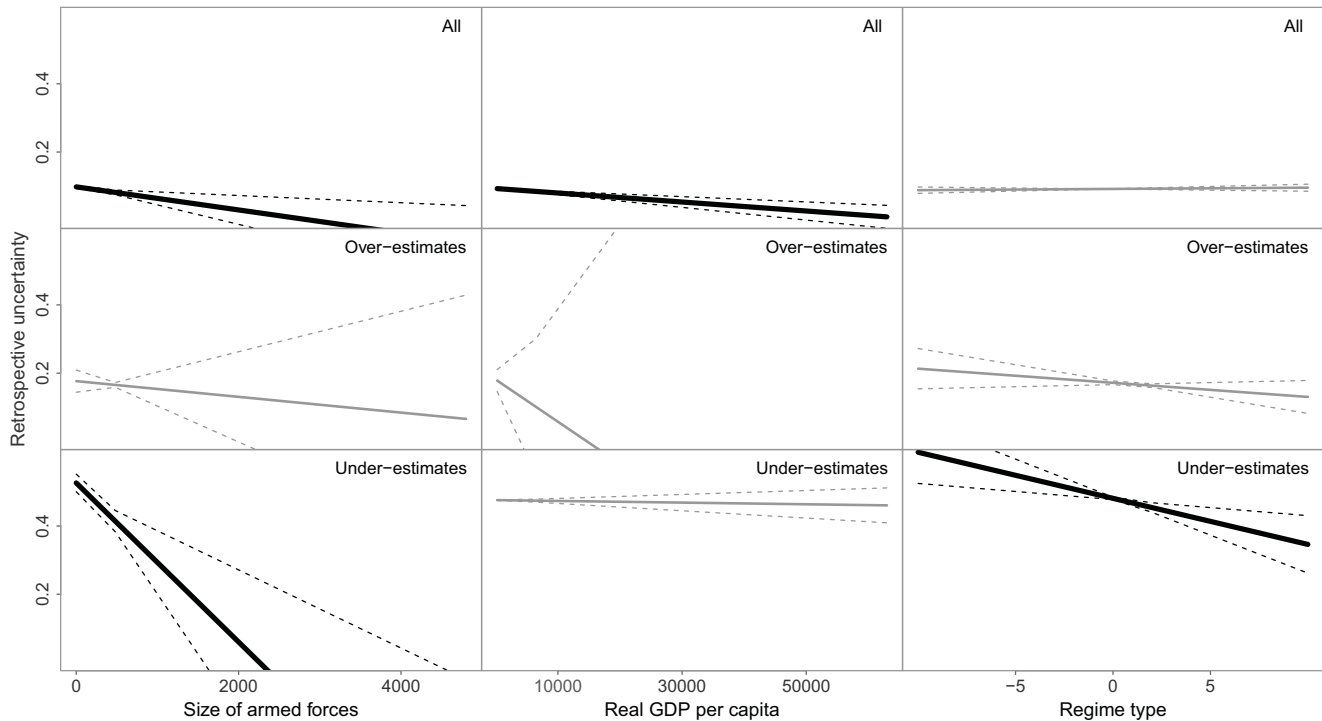


Figure 2. Substantive effects on retrospective uncertainty.

within-country results are used in the figure. Black lines represent statistically significant relationships, while gray lines fail to reach significance at the 95 percent confidence level; the dashed lines are 95 percent confidence intervals.

For coder-assessed uncertainty, too, our results are substantively significant. When comparing a state with the largest armed forces in our data to a state with the smallest, while holding other variables at their mean, the maximum number of military personnel is associated with an 81 percentage point greater likelihood that the estimate will be coded as particularly uncertain. Shifting a state from the minimum to the maximum Polity score is associated with about a 1.2 percentage point increase in the probability that coders will assess the estimate as uncertain, holding other variables at their means. Within-country effects for international alliances (2.8 percentage points), conflict history (0.5 percentage point), and nuclear status (3.1 percentage points) are of the same magnitude.

The size of this on-average substantive effect, while significant in the context of uncertainty over military capabilities, appears low largely because coder assessments of uncertainty are relatively unusual in our data. Most country estimates have very little risk of being seen as particularly uncertain. A perhaps more useful way to understand substantive effects in this case is to evaluate the impact of variables of interest among the country-years that have at least some chance of uncertain estimates—these are, after all, the observations of most interest for the study of uncertainty in international assessments (Beck, King, and Zeng 2000; Kaplow 2016).

Figure 3 shows the change in the predicted probability of coder-assessed uncertainty when a particular factor is shifted from two standard deviations below its mean to two standard deviations above its mean.²¹ Each horizontal bar represents a 95 percent confidence interval; dark bars show factors that

are statistically significant in the underlying model. These first differences are calculated not by holding all other variables at their global mean, but rather by holding other variables at the mean among those observations predicted by the model to have greater than 5 percent chance of being coded as uncertain. Figure 3, then, shows the substantive effect on coder-assessed uncertainty for cases that have some significant risk of being judged uncertain. These results are more extreme than those calculated using the global mean, but they represent a more relevant picture of the effect of these variables in cases of particular interest.

Conclusion

Uncertainty has been theorized to play an important role in a variety of international relations phenomena, from conflict, to trade disputes, to compliance with international agreements. However, scholars have struggled to obtain empirical purchase on the concept. In this article, we use new measures of uncertainty over military capabilities to examine what factors lead to accurate international assessments or those in which states are most confident.

We find that both the characteristics of states and the elements of their strategic environment have a strong and significant effect on uncertainty, but the role of these factors depends on the context of the assessment. International assessments are likely to become more accurate as target states increase their wealth, the number of personnel in their armed forces, or their level of democracy, while nuclear acquisition is associated with higher—but not necessarily more accurate—estimates. The same factors, however, seem to have a contrasting effect on analysts' confidence in their international assessments. Coders are more likely to assess as uncertain states with larger militaries, more democratic regimes, international alliances, or involvement in recent disputes.

²¹ Calculations in figure 3 are based on Model 4.

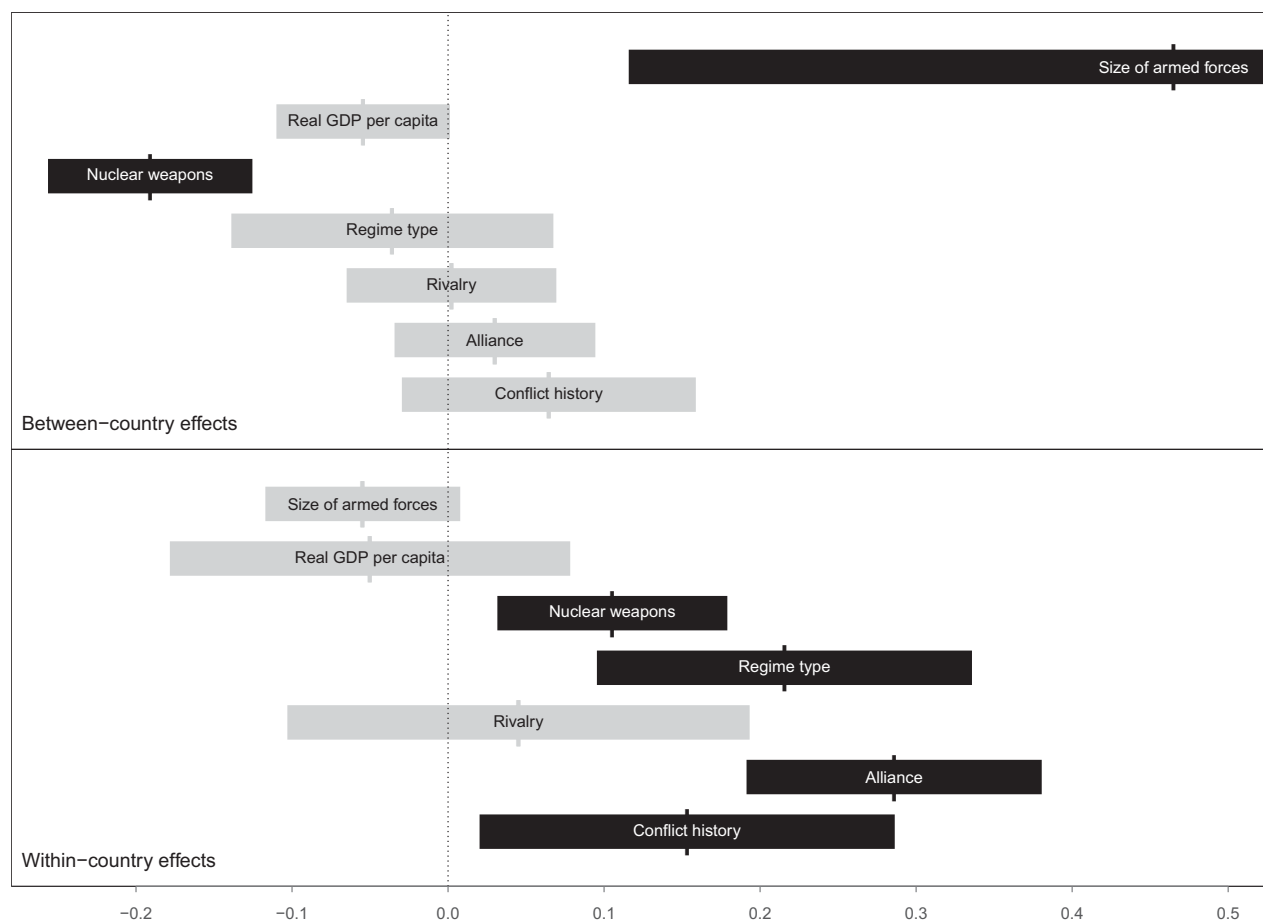


Figure 3. Substantive effects on coder-assessed uncertainty.

While we have offered some initial hypotheses, several of our findings suggest the need for additional research. Why do alliance ties and conflict history seem to have such a limited effect on uncertainty measured retrospectively? Do nuclear weapons lead to overestimates because of the added complexity of the target, changing perceptions of the now-nuclear state, or some other mechanism? Why is the confidence states have in their assessments more sensitive to the strategic environment than the retrospective accuracy of those assessments?

More broadly, this work provides a proof of concept for a general approach to measuring uncertainty in a variety of international domains, exploiting estimates of state characteristics that overlap or change over time. Our method is general in that it can be applied to questions in international relations across substantive areas, and scalable in that it can be extended to take advantage of new data as these become available.

These measures of uncertainty also allow scholars and analysts to distinguish between two types of uncertainty—“known unknowns” and “unknown unknowns.” This distinction opens a variety of new avenues for theory building in future research: Do states correctly assess their own level of uncertainty? Can states do a better job of mitigating the negative effects of uncertainty when they are aware of the potential for misperception? Does the knowledge of others’ uncertainty allow states to strategically manipulate this uncertainty in others to their own advantage?

Finally, uncertainty is not a static attribute of the international system. States and other actors have a variety of tools

at their disposal to build up or diminish uncertainty about capabilities and resolve, ranging from benign government reports to flat-out denial and deception against enemy intelligence capabilities. Future research might fruitfully examine the mechanisms by which states can manipulate uncertainty, identifying those situations in which adversaries are more or less likely to conceal military strength or weakness. Understanding these dynamics would then allow for a more clear-eyed assessment of state capabilities and their likely impact on international conflict.

Supplementary Information

Supplementary information is available at the *International Studies Quarterly* data archive.

References

- ANDERTON, CHARLES H. 1989. “Arms Race Modeling: Problems and Prospects.” *Journal of Conflict Resolution* 33 (2): 346–67.
- BARTELS, BRANDON. 2015. “Beyond ‘Fixed versus Random Effects’: A Framework for Improving Substantive and Statistical Analysis of Panel, Time-Series Cross-Sectional, and Multilevel Data.” In *Quantitative Research in Political Science*, edited by Robert J. Franzese. Los Angeles, CA: Sage.
- BEARCE, DAVID H., KRISTEN M. FLANAGAN, AND KATHARINE M. FLOROS. 2006. “Alliances, Internal Information, and Military Conflict Among Member-States.” *International Organization* 60 (3): 595–625.
- BECK, NATHANIEL, AND JONATHAN N. KATZ. 1996. “Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models.” *Political Analysis* 6 (1): 1–36.

- _____. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science* 14 (1): 331–52.
- BECK, NATHANIEL, GARY KING, AND LANGCHE ZENG. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94 (1): 21–35.
- BELL, ANDREW, AND KELVYN JONES. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3 (1): 133–53.
- BENNETT, D. SCOTT, AND ALLAN C. STAM. 2004. *The Behavioral Origins of War*. Ann Arbor, MI: University of Michigan Press.
- BLAINEY, GEOFFREY. 1973. *The Causes of War*. New York: Free Press.
- BOEHMER, CHARLES, GARTZKE ERIK, AND NORDSTROM TIMOTHY. 2004. "Do Intergovernmental Organizations Promote Peace?" *World Politics* 57 (1): 1–38.
- BRZOSKA, MICHAEL. 1981. "The Reporting of Military Expenditures." *Journal of Peace Research* 18 (3): 261–75.
- _____. 1995. "World Military Expenditures." In *Handbook of Defense Economics*, edited by Keith Hartley and Todd Sandler. Amsterdam: Elsevier, 45–67.
- CARTER, DAVID B., AND CURTIS S. SIGNORINO. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18 (3): 271–92.
- DEUTSCH, KARL W., AND J. DAVID SINGER. 1964. "Multipolar Power Systems and International Stability." *World Politics* 16 (3): 390–406.
- FEARON, JAMES D. 1995. "Rationalist Explanations for War." *International Organization* 49 (3): 379–414.
- _____. 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41 (1): 68–90.
- FENG, YI. 2001. "Political Freedom, Political Instability, and Policy Uncertainty: A Study of Political Institutions and Private Investment in Developing Countries." *International Studies Quarterly* 45 (2): 271–94.
- FITZPATRICK, MARK. 2011. *North Korean Security Challenges: A Net Assessment*. London: International Institute for Strategic Studies.
- GARTHOFF, RAYMOND L. 2004. "Estimating Soviet Military Intentions and Capabilities." In *Watching the Bear: Essays on CIA's Analysis of the Soviet Union*, edited by Gerald K. Haines and Robert E. Leggett. Washington, DC: Center for the Study of Intelligence.
- GARTZKE, ERIK. 1999. "War Is in the Error Term." *International Organization* 53 (3): 567–87.
- _____. 2013. "The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth." *International Security* 38 (2): 41–73.
- GARTZKE, ERIK, JEFFREY M. KAPLOW, AND RUPAL N. MEHTA. 2014. "The Determinants of Nuclear Force Structure." *Journal of Conflict Resolution* 58 (3): 481–508.
- GARTZKE, ERIK, AND MATTHEW KROENIG. 2009. "A Strategic Approach to Nuclear Proliferation." *Journal of Conflict Resolution* 53 (2): 151–60.
- GHOSN, FATEN, GLENN PALMER, AND STUART A. BREMER. 2004. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21 (2): 133–54.
- GIBLER, DOUGLAS M., AND MEREDITH REID SARKEES. 2004. "Measuring Alliances: The Correlates of War Formal Interstate Alliance Dataset, 1816–2000." *Journal of Peace Research* 41 (2): 211–22.
- GLASER, CHARLES L. 1997. "The Security Dilemma Revisited." *World Politics* 50 (1): 171–201.
- GLEDITSCH, KRISTIAN SKREDE. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46 (5): 712–24.
- HOLMES, MARCUS. 2013. "The Force of Face-to-Face Diplomacy: Mirror Neurons and the Problem of Intentions." *International Organization* 67 (4): 829–61.
- HOROWITZ, MICHAEL C., AND ALLAN C. STAM. 2014. "How Prior Military Experience Influences the Future Militarized Behavior of Leaders." *International Organization* 68 (3): 527–59.
- HUTH, PAUL, D. SCOTT BENNETT, AND CHRISTOPHER GELPI. 1992. "System Uncertainty, Risk Propensity, and International Conflict among the Great Powers." *Journal of Conflict Resolution* 36 (3): 478–517.
- JERVIS, ROBERT. 1976. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.
- _____. 1978. "Cooperation under the Security Dilemma." *World Politics* 30 (2): 167–214.
- KAPLOW, JEFFREY M. 2016. "The Negotiation Calculus: Why Parties to Civil Conflict Refuse to Talk." *International Studies Quarterly* 60 (1): 38–46.
- KYDD, ANDREW. 2003. "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47 (4): 597–611.
- LEBLANG, DAVID. 2003. "To Devalue or to Defend? The Political Economy of Exchange Rate Policy." *International Studies Quarterly* 47 (4): 533–60.
- LEBOVIC, JAMES H. 1995. "How Organizations Learn: U.S. Government Estimates of Foreign Military Spending." *American Journal of Political Science* 39 (4): 835–63.
- _____. 1998. "Consider the Source: Organizational Bias in Estimates of Foreign Military Spending." *International Studies Quarterly* 42 (1): 161–74.
- _____. 1999. "Using Military Spending Data: The Complexity of Simple Inference." *Journal of Peace Research* 36 (6): 681–97.
- MARSHALL, MONTY G., KEITH JAGGERS, AND TED ROBERT GURR. 2010. *Polity IV Project: Political Regime Characteristics and Transitions, 1800–2010*. Vienna, VA: Center for Systemic Peace.
- MEIROWITZ, ADAM, AND ANNE E. SARTORI. 2008. "Strategic Uncertainty as a Cause of War." *Quarterly Journal of Political Science* 3 (4): 327–52.
- MELLERS, BARBARA, E. STONE, P. ATANASOV, N. ROHRBAUGH, S.E. METZ, L. UNGAR, M.M. BISHOP, M. HOROWITZ, E. MERKLE, AND P. TETLOCK. 2015. "The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics." *Journal of Experimental Psychology: Applied* 21 (1): 1–14.
- MORROW, JAMES D. 1989. "Capabilities, Uncertainty, and Resolve: A Limited Information Model of Crisis Bargaining." *American Journal of Political Science* 33 (4): 941–72.
- NOREN, JAMES. 2003. "CIA's Analysis of the Soviet Economy." In *Watching the Bear: Essays on CIA's Analysis of the Soviet Union*, edited by Gerald K. Haines and Robert E. Leggett. Washington, DC: Center for the Study of Intelligence.
- PRADOS, JOHN. 1982. *The Soviet Estimate: U.S. Intelligence Analysis & Russian Military Strength*. New York: Dial Press.
- REED, WILLIAM. 2003a. "Information and Economic Interdependence." *Journal of Conflict Resolution* 47 (1): 54–71.
- _____. 2003b. "Information, Power, and War." *American Political Science Review* 97 (4): 633–41.
- RIDER, TOBY J. 2013. "Uncertainty, Salient Stakes, and the Causes of Conventional Arms Races." *International Studies Quarterly* 57 (3): 580–91.
- SECHSER, TODD S., AND ELIZABETH N. SAUNDERS. 2010. "The Army You Have: The Determinants of Military Mechanization, 1979–2001." *International Studies Quarterly* 54 (2): 481–511.
- SHIRKEY, ZACHARY C. 2016. "Uncertainty and War Duration." *International Studies Review* 18 (2): 244–67.
- SINGER, J. DAVID. 1988. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985." *International Interactions* 14 (2): 115–32.
- SINGER, J. DAVID, STUART BREMER, AND JOHN STUCKEY. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820–1965." In *Peace, War, and Numbers*, edited by Bruce M. Russett. Beverly Hills, CA: Sage.
- SLANTCHEV, BRANISLAV L. 2003. "The Principle of Convergence in Wartime Negotiations." *American Political Science Review* 97 (4): 621–32.
- _____. 2004. "How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace." *American Journal of Political Science* 48 (4): 813–29.
- _____. 2010. "Feigning Weakness." *International Organization* 64 (3): 357–88.
- SMITH, ALASTAIR, AND ALLAN C. STAM. 2004. "Bargaining and the Nature of War." *Journal of Conflict Resolution* 48 (6): 783–813.
- SMITH, BRADLEY C., AND WILLIAM SPANIEL. 2019. "Militarized Disputes, Uncertainty, and Leader Tenure." *Journal of Conflict Resolution* 63 (5): 1222–52.
- THOMPSON, WILLIAM R., AND DAVID R. DREYER. 2011. *Handbook of International Rivalries: 1494–2010*. Washington, DC: Congressional Quarterly Press.
- TREVERTON, GREGORY F. 2003. *Reshaping National Intelligence for an Age of Information*. Cambridge: Cambridge University Press.
- US ARMS CONTROL AND DISARMAMENT AGENCY. 1966–1997. *World Military Expenditures and Arms Transfers*. Washington, DC.
- US DEPARTMENT OF STATE. 1998–2019. *World Military Expenditures and Arms Transfers*. Washington, DC.
- WAGNER, R. HARRISON. 2000. "Bargaining and War." *American Journal of Political Science* 44 (3): 469–84.
- WALTZ, KENNETH NEAL. 1979. *Theory of International Politics*. Long Grove, IL: Waveland Press.
- WOLFORD, SCOTT. 2007. "The Turnover Trap: New Leaders, Reputation, and International Conflict." *American Journal of Political Science* 51 (4): 772–88.
- YARHI-MILO, KEREN. 2013. "In the Eye of the Beholder: How Leaders and Intelligence Communities Assess the Intentions of Adversaries." *International Security* 38 (1): 7–51.
- _____. 2014. *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations*. Princeton, NJ: Princeton University Press.