



A production of H-Diplo with the journals *Security Studies*, *International Security*, *Journal of Strategic Studies*, and the International Studies Association's Security Studies Section (ISSS).

<http://issforum.org>

H-Diplo/ISSF Editors: **Thomas Maddux and Diane Labrosse**

H-Diplo/ISSF Web and Production Editor: **George Fujii**

Commissioned for H-Diplo/ISSF by **Thomas Maddux**

---

## H-Diplo | ISSF Article Review 56

**Aaron Rapport.** "Hard Thinking about Hard and Easy Cases in Security Studies." *Security Studies* 24:3 (July-September 2015): 431-465. DOI: 10.1080/09636412.2015.1070615.

<http://dx.doi.org/10.1080/09636412.2015.1070615>

Review by **Timothy McKeown**, University of North Carolina, Chapel Hill

Published by ISSF on 2 June 2016

[tiny.cc/ISSF-AR56](http://tiny.cc/ISSF-AR56)

<https://issforum.org/articlereviews/56-hard-thinking>

<https://issforum.org/ISSF/PDF/ISSF-AR56.pdf>

---

American political scientists who perform case studies straddle two very different strategies of inquiry. On the one hand, they seek to understand politics in terms of narratives that explore closely the events in a single historical episode. On the other, they work within a disciplinary framework that emphasizes the statement of general theories, the treatment of events as samples drawn from a larger population, and hypothesis testing using statistical inference as a central activity. That these two perspectives are deeply in tension has been recognized since quantitative social science began in the early twentieth century, and has been the source of seemingly never-ending disciplinary controversy.<sup>1</sup>

---

<sup>1</sup> Martyn Hammersley, *The Dilemma of Qualitative Method: Herbert Blumer and the Chicago Tradition* (New York: Routledge, 1989).

The current era has been marked by a number of attempts to reconcile these two perspectives, primarily by borrowing concepts from statistical hypothesis testing and translating them into operations in case studies.<sup>2</sup> Aaron Rapport's recent article is an example of this trend. He attempts to appropriate the logic of Bayesian statistical inference to explicate the logic of case selection – specifically, the idea of focusing on “most likely” or “least likely” cases because they seemingly provide the largest relative improvements of our assessments of the empirical accuracy of the relevant theories.

The terminology of ‘most likely’ and ‘least likely’ cases refers to researchers’ initial assessments of events. A case is ‘least likely’ when, from the standpoint of a given theory, the initial conditions in the case imply that the observed process or outcome in the case is quite unlikely – indeed, literally less likely than in any other set of initial conditions that might apply to this class of cases. Likewise, ‘most likely’ cases are ones where the initial conditions imply that a given outcome is the most probable. (Because these requirements are stringent, much of Rapport's discussion involves ‘more’ and ‘less’ likely cases).

Rapport contends that ‘most likely’ and ‘least likely’ cases are attractive because they give researchers a better chance of using case-study findings to make judgements about the theories that account for events in the underlying population from which the case study is drawn. What he means by that can be illustrated by the following example using Bayes's rule for revising probability estimates to analyze theory and evidence relating to war involvement:

Let

$p(\text{Th})$  = the probability that some theory about war involvement is correct

$p(w)$  = the unconditional probability that a nation is involved in war

$p(w|\text{Th})$  = the conditional probability that a nation in a specific situation will be involved in a war, given the predictions of the theory for that situation

$p(\text{Th}|w)$  = the probability that the theory is correct, given that the nation was in a war

Bayes's rule is derived from the definition of a conditional probability. It states that

$$p(\text{Th}|w) = (p(\text{Th})/p(w)) * p(w|\text{Th})$$

If one begins with some probability  $p(\text{Th})$  that a theory is true, Bayes's rule tells us how to use information about subsequent events (here, a war) to update that probability.

A ‘least likely’ case would be one where  $p(w|\text{Th})$  were at its minimum. If this term were zero, then the right-hand side of the equation would equal zero, and an observation of war would mean that  $p(\text{Th}|w)$  must also be zero. Thus, the theory would be rejected, regardless of the unconditional probability of war  $p(w)$  or the researcher's degree of initial confidence in the theory  $p(\text{Th})$ . Where  $p(w|\text{Th})$  is greater than zero, the posterior assessment of the probability that the theory is correct will be greater as prior confidence in the theory  $p(\text{Th})$

---

<sup>2</sup> Many of these works are cited in Rapport's article.

is greater, as the unconditional probability of war involvement  $p(w)$  declines, and as the theory assigns a higher probability to war via the  $p(w|Th)$  term. As long as they share the same definition of war, researchers will find it relatively easy to agree on the unconditional probability of war involvement. They are much less likely to agree on how much confidence to attach to the truth of an existing theory, or on which case or cases are minimum probability cases for the occurrence of war, or what that probability is. If they cannot agree on these, then it is no wonder that they cannot agree on the posterior assessment of the theory, given the results of the case study.

Rapport claims to provide "concrete steps ... to determine how hard a case is for particular theories" (432). This seems to be advice on how to set  $p(w|Th)$ , or to find cases that correspond to minima or maxima for it. The long discussion of 'countervailing conditions' and 'Bayesian' approaches seems to be providing advice on setting  $p(Th)$ ; the former approach focuses on an assessment peculiar to the case under consideration, while the latter suggests taking into account more information. Why one of these is more 'Bayesian' than the other isn't clear.

Although this discussion is comprehensible from a Bayesian perspective, it is preoccupied with concerns that Bayesian statistical theory does not formally address. Bayesians treat probability assessments as subjective, resting on a complex mix of experience, evidence, and theory. The theory does not inquire into how people do or should form them, focusing instead on how to use new information to update them. Prior beliefs merely initialize the updating process, and the influence of the prior beliefs dwindles as new observations accumulate.

The world of case-study researchers is typically quite different. Cases might be quite scarce (for example, there are very few instances of modern great power wars, hence a universe of such cases will be too small for meaningful statistical analysis). With very small universes, the influence of prior assessments will not fade away. Thus, prior assessments are far more consequential than they ordinarily are in statistical analysis.

Although Rapport frames his discussion in terms of presenting ways to perform these assessments, the practical advice that he offers leaves this reader pessimistic that much can be accomplished. Indeed, Rapport's own conclusion is that "Given the limitations of theory, as well as uncertainty involved with measurement and interpretation of evidence, claims that a certain case represents a hard or easy test can normally be challenged on reasonable grounds" (464). He presents several reasons. First, the theory under examination has to "posit clear probability relationships or determinative laws" (435). With theories that state relationships only verbally, probability assessments are at best ordinal – something is perhaps more or less likely than something else, but it is not obvious that this is what Rapport means by a "clear" assessment. Second, the difference between the observed outcome and the theoretically expected outcome is treated as distance in a Euclidian space (437). For any variable that varies continuously, the absence of a measure of distance means that these judgments are extremely weak. When two case studies both have findings at variance from expected results, how would we decide which study comes closer to what has been predicted? Likewise, how can a case study be used to infer "how strongly the variables that [the theory] highlights impact the dependent variable [i.e. the outcome]" (444)? Rapport mentions only that such judgments can be based on the results of prior research (445), but this does not clarify whether the case study itself provides a basis for updating these assessments. Third, the case selected should be one where theory predicts an extreme outcome (438). Again, if the scaling of outcomes is only ordinal, then we can distinguish between "a lot" and "a little," and perhaps rank-order a very small set of outcomes from least-likely to most-likely, but only in that limited sense can we identify outcomes as extreme. Fourth, the theory must be simple enough that an expected outcome can be

identified (442-443). When a variety of explanatory factors are considered, and they are posited to have conflicting effects on the outcome, then the absence of weights for these factors means that the expected outcome in the case is indeterminate.

If the only purpose of case studies is to serve as the test of a theory, then one might wonder whether they are worth the trouble. Neither Rapport nor the literature that he cites offers a basis for more than very modest expectations about what can be accomplished. Using large data sets and statistical procedures to test theories seems much less problematic.

The comparative advantage of case studies lies not in testing theories, but in uncovering the process whereby outcomes are created. Rapport notes that theories that “carefully specify the intervening mechanisms” between initial conditions and observed responses clarify how outcomes are generated (444). They also clarify how existing theories fail to account for the flow of events. This is not a statistical test, but it is valuable in reformulating theories and developing new ones. Rapport emphasizes that a theory that makes predictions about the process can be tested several times in a single case (434, 444). (To continue the statistical analogy, this is akin to time-series analysis, where a sequence of outcomes is predicted and then compared to observed events.) An obvious difficulty is that we have only chunks of theory that partially accomplish this task (cognitive and social psychology, theories of networks and organizations, and public choice theories of two-level games or structure-induced equilibria), and none of them has much to do with realist or neo-realist theory. Indeed, the realist or neo-realist foundational assumption is that external situations compel rational governments to respond in certain ways, or else face their own demise. The more seriously one takes that foundational assumption, the more one is driven to Kenneth Waltz’s position that realist theory is about selection mechanisms, not decision processes.<sup>3</sup> One can of course ‘fill in’ realist theories by adopting various auxiliary arguments that depict the decision-making process, but this is a tacit admission that a theory that relies solely on situational determinism is inadequate for understanding foreign policy-making.<sup>4</sup>

The situation with regard to constructivist and public-choice (i.e. ‘liberal’) theories is similar, as they also ‘fail to specify the intervening mechanisms’ and researchers must add various auxiliary claims if they are to connect the theory to an observed decision process. The result is a profusion of case-specific renderings of these theories, which predictably produces the cacophony of contending positions found in Rapport’s discussion.

What then is accomplished by attempts to develop the logic of most likely versus least likely case selection rules? One achievement is that the framework for case selection and assessment provided by Bayes’s rule clarifies why researchers gravitate to cases that seem puzzling or anomalous. Cases where expectations are strongly violated have the potential to lead to relatively larger revisions in assessments of the truth of

---

<sup>3</sup> Kenneth N. Waltz, “International Politics is not Foreign Policy,” *Security Studies* 6.1 (1996): 54-57.

<sup>4</sup> A longer discussion of realist theory and selection is in Timothy J. McKeown, “The Limitations of ‘Structural’ Theories of Commercial Policy,” *International Organization* 40.1 (1986): 43-64. The literature on state deaths as a selection mechanism is briefly reviewed in Timothy J. McKeown, “Neorealism,” in *Oxford Bibliographies – International Relations*, ed. David Armstrong (New York: Oxford University Press, 2014). Especially noteworthy is Dustin E. Howes, “When States Choose to Die: Reassessing Assumptions about What States Want,” *International Studies Quarterly* 47.4 (2003): 669-692.

theoretical claims – thus some cases are rightly seen as more critical for theory assessment than others.<sup>5</sup> Another clarification provided by Bayesian thinking is a more explicit understanding of how prior understandings shape the assessment of research results, and how differences in understandings seldom disappear quickly in the case study world. For Bayesians, individuals converge on their estimates of the truth of various claims only as evidence accumulates, and the impact of initial differences in prior assessments gradually weakens. If cases cumulate very slowly, convergence will also be very slow.

The effort to use case studies to test structural theories such as realism brings to light difficulties that are bound to arise when an empirical procedure whose chief advantage is its ability to uncover and document social processes is used to test a theory that relies on an input-output logic and deliberately neglects social processes. It seems unlikely that an adequate theory of foreign policy outcomes can be constructed without developing an accurate and detailed understanding of the social processes that generate them. Theories that directly address these processes seem more likely to lead to productive results than those that can only address processes by means of various auxiliary mechanisms tacked onto to their original structure. An emphasis on hypothesis testing seems to presuppose that we are blessed with an abundance of plausible theories, and our main task is to weed out the weaker candidates. An empirical procedure that aids theory construction or revision seems at least as important. Case studies will continue to be performed by political scientists to aid in theory development, regardless of how well or how poorly the cases test theories.<sup>6</sup> The rich accounts of the decision process produced by diplomatic historians directly contribute to theory development in this way.

**Timothy J. McKeown** is Professor of Political Science at the University of North Carolina, Chapel Hill. His research focuses on international political economy and foreign policy. His most recent publication is “A different two-level game: Foreign policy officials’ personal networks and policy coordination,” *Review of International Political Economy* 23.1 (2016): 93-122. He is currently working on long-run changes in the use of realist policy justifications by U.S. foreign policy officials.

**Copyright ©2016 The Authors.**

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 United States License](https://creativecommons.org/licenses/by-nc-nd/3.0/)

---

<sup>5</sup> Bayesian statisticians who have commented on the logic of case selection stress the importance of anomalous cases as spurs to theory development. Andrew Gelman and Thomas Basbøll, “When Do Stories Work? Evidence and Illustration in the Social Sciences,” *Sociological Methods & Research* 43.4 (2014): 547-570.

<sup>6</sup> These points are discussed at greater length in McKeown, “The Limitations of ‘Structural’ Theories” and idem, “Case Studies and the Statistical Worldview,” *International Organization* 53.1 (1999): 161-190. For process-oriented theories that are directly helpful in understanding foreign policy decision-making, see the literature cited in idem, “A Different Two-Level Game: Foreign Policy Officials’ Personal Networks and Policy Coordination,” *Review of International Political Economy* 23.1 (2016): 93-122.