Hard Thinking about Hard and Easy Cases in Security Studies

Aaron Rapport
University of Cambridge
ar727@cam.ac.uk
Forthcoming, *Security Studies*

Even scholars who believe the case study is a valuable research tool are skeptical that researchers can effectively generalize case findings to a wider population.[1] However, an exception may be possible with "least likely" (LL) and "most likely" (ML) cases, often referred to as "hard" and "easy" cases, respectively. The former pose "difficult" tests of theories, in that—unlike ML cases—one would not expect a theory's expectations to be borne out by a review of the case evidence. According to Eckstein, LL and ML cases provide researchers with a great deal of inferential leverage, being "especially tailored" to either confirming or disconfirming a theory, respectively.[2] Such case study designs are common in the discipline of International Relations, particularly in the subfield of security studies. Since 1997, at least 36 articles appearing in *Security Studies* explicitly invoked the logic of LL/ML studies to justify their case selection and defend or impugn the validity of particular theories.[3] The appeal of a study that might offer in-depth knowledge of a particular case and the generalizability of large-n cross-unit comparisons is obvious.

Despite the widespread use of LL/ML cases there remains a considerable amount of imprecision regarding their definition. Researchers' applications of the LL or ML labels typically appear plausible, but further investigation often yields uncertainty about the validity of a case's purported difficulty. Such questions have been central to some of the most significant developments in security studies in the past two decades. For instance, the emergence of constructivist scholarship on "traditional" security issues in the 1990s had a major impact on the field, in part because such matters were alleged to be particularly hard cases for theories that

---

[1] John Gerring, "What is a Case Study and What is it Good for?" *American Political Science Review* 98, no. 2 (2004): 347.

[2] Harry Eckstein, "Case Study and Theory in Political Science," in *Handbook of Political Science, Vol. 7: Strategies of Inquiry*, eds. Fred I. Greenstein and Nelson W. Polsby (Reading, MA: Addison-Wesley),118-119.

[3] I searched the journal's electronically-accessible articles using eight phrases: "least likely case"; "most likely case"; "hard case"; "easy case"; "hard test"; "easy test"; "most difficult case"; and "least difficult case." In the same time period, these search parameters found 18 articles that used an LL/ML case selection strategy from *International Organization*; 19 from *International Security*; and 17 from *International Studies Quarterly*.

privileged norms, identity, and culture to explain.[4] However, this assertion was soon challenged by scholars who contended that constructivists studying international security had actually chosen easy cases to test their theories.[5] This and other episodes demonstrate that the logic surrounding LL/ML cases deserves careful scrutiny, both for the sake of methodological rigor and disciplinary progress.

This article addresses ambiguities surrounding LL/ML cases and provides concrete steps researchers are encouraged to follow to determine how "hard" a case is for particular theories. It uses examples from research on security matters to demonstrate how scholars can draw generalizable inferences from cases, as well as illustrate potential pitfalls researchers may confront. The first section of this article introduces the two ways scholars typically justify claims that a case is LL or ML, what I deem the "countervailing conditions" and Bayesian approaches. It briefly outlines the relationship between the two approaches, particularly focusing on the consequences of neglecting Bayesian insights when drawing inferences from case evidence. The next two sections more closely examine the core elements of the CC and Bayesian rationales, respectively, further explaining the assertions made in the first section. The necessary steps a researcher must take to determine the degree to which a case is ML or LL are also laid out in detail. The fourth section examines a study of the non-use of nuclear weapons during the Vietnam War to show how better understanding of both rationales can affect how cases are categorized and inferences reached. The article concludes by considering how LL/ML cases can contribute to research in security studies given constraints imposed by theory and evidence.

---

[4] Peter J. Katzenstein, "Introduction: Alternative Perspectives on National Security," in *The Culture of National Security: Norms and Identity in World Politics*, ed. Peter J. Katzenstein (New York: Columbia University Press, 1996), 11.

[5] Michael C. Desch, "Culture Clash: Assessing the Importance of Ideas in Security Studies," *International Security* 23, no. 1 (1998): 141-170.

RATIONALES FOR CHOOSING LL AND ML CASES

Researchers' rationales for describing cases as LL or ML generally fall into two categories. The most common is the countervailing conditions (CC) approach, which focuses on the details of a specific case. According to this rationale, LL cases are those with characteristics the researcher expects will promote some outcome *other* than that predicted by the primary theory they are testing. Conversely, ML cases feature characteristics that should facilitate rather than hinder the outcome expected by said theory. The second, increasingly discussed rationale used to identify LL/ML cases follows a Bayesian logic. This approach attends more to the population from which a case was drawn, defining a case as LL or ML according to a researcher's prior confidence that the theory being tested offers a valid explanation for the outcome of interest in other, similar cases. Regardless of rationale, cases fall on a continuum of how likely they are to support a theory's expectations.[6] Cases are apt to be "less" or "more" likely to support a theory rather than "least" or "most." For simplicity, however, this article refers to cases as if they are LL, ML, or neither.

As will be shown, the CC approach is actually nested within a Bayesian framework, and researchers who ignore the implications of Bayesian reasoning may draw substantially different inferences from their case evidence than if they had taken prior confidence in a theory into account. First, if prior confidence in a novel theory is neglected, researchers will be more liable to define cases used to test the theory as ML. This in turn increases the chances they will reject theories whose expectations are not met in initial case studies. Alternatively, Bayesianism lowers one's confidence that any case will support the expectations of a novel theory, given the lack of prior information regarding the theory's explanatory power and scope conditions needed for it to

[6] Alexander L. George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge, MA: MIT Press, 2004), 122.

operate. Researchers who factor Bayesian insights into their studies should thus be more willing to subject new theories to further tests even if they "fail" in an initial case, selecting new cases based on updated beliefs about where the theory might be applied. Second, according to the Bayesian perspective the difficulty of a case depends in part on the number of theories that could explain the outcome of interest. Every plausible alternative to the theory of interest the researcher identifies will decrease confidence in the latter, even if all the theories expect the same outcome in a case. Researchers who neglect Bayesianism will thus again be more likely to apply the ML label to cases and firmly reject theories that fail to account for case outcomes. Lastly, from the Bayesian perspective there is such a thing as cases that are too hard or easy in that they do not allow researchers to generalize about the theory of interest. Overly ML cases are those that resemble numerous others that a theory has convincingly explained in the past. Because confidence in the theory in question is so high to begin with, a single "failure" in such a case will not do much to undermine that confidence.[7] Similarly, overly LL cases resemble others which a theory has repeatedly failed to explain in the past, or those an alternative theory has explained in an overwhelmingly convincing fashion. In contrast, those who neglect prior confidence in a theory presume that the harder or easier a case is the more inferential leverage it provides.

THE COUNTERVAILING CONDITIONS RATIONALE

A countervailing condition is a variable whose presence (or specific value) in a chosen case decreases the probability the researcher will observe the outcome posited by the theory

[7] A finding that is anomalous solely in light of theoretical expectations, rather than theoretical expectations as well as established empirical findings, is more indicative that researchers should generate new explanations. See Ronald Rogowski, "The Role of Theory and Anomaly in Social-Scientific Inference," *American Political Science Review* 89, no. 2 (1995): 467-470. As noted, however, an anomalous finding need not lead one to forgo additional tests of a novel theory.

being tested. Though he did not use the terminology here, Eckstein expressed the CC rationale for designating cases as LL/ML in his seminal article on case study methods.[8] Eckstein did not lay out precise steps for determining how difficult a test a case posed for a particular theory, instead using Michels' proposition that oligarchy is an "iron law" in organizations to demonstrate the LL case-selection strategy.[9] Michels chose to examine horizontally structured political parties with strongly expressed democratic values. He reasoned that if rule by small groups of elites was observed in such cases, more confidence could be placed in his assertions about the ubiquity of oligarchy. In sum, Michels chose cases in which he determined there were strong countervailing conditions that would hinder the emergence of oligarchy, oligarchic rule being the outcome he expected to observe given his own theory.

The CC logic is often invoked by researchers in security studies whose theories stress the importance of values, norms, and ideas rather than states' material capacities to explain case outcomes. A typical version of this argument holds that "Because realists argue that ideational factors are even less important when national security (or 'high politics') is at stake, ideational variables should exert little influence [on policy choice]."[10] In other words, the high stakes realists assert are attached to security politics are expected to overwhelm evaluations of whether a policy option is consistent with an actor's social identity or notions of appropriateness, even if these latter considerations are politically salient. If this is so, "ideational variables" would have to exert especially strong effects on policy choice for their expected outcome to prevail when realism predicted a divergent outcome in the same case.

---

[8] Eckstein, "Case Study and Theory," 116-119.

[9] Robert Michels, *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy* (New York: Collier Books, 1962).

[10] Giorgi Gvalia, David Siroky, Bidzina Lebanidze, and Zurab Iashvili, "Thinking Outside the Bloc: Explaining the Foreign Policies of Small States," *Security Studies* 22, no. 1 (2013): 100.

There are three general steps necessary for selecting a case using the CC rationale. These steps may appear misleading in their simplicity; complications in the CC case-selection process are addressed in the subsections that follow. First, researchers must make sure to select cases that fall within the *scope conditions* of both the theory of interest—the primary theory being tested—and those of any alternative explanations under examination. Second, for all theories under examination the researcher must be able to derive *clear, specific expectations* about what outcome should be observed given the antecedent conditions in the chosen case. Cases cannot pose LL/ML tests unless the theories under examination posit clear probabilistic relationships or determinative laws.[11] They should be precise enough that, for categorical dependent variables, only one outcome can be consistent with a theory's expectations in a given case, or a small range of outcomes for continuous dependent variables. If numerous alternative outcomes in a case can be reconciled with a theory post hoc, evaluating the fit between theory and evidence becomes difficult. It is also highly desirable to be able to clearly specify the intervening mechanisms between the explanatory variables highlighted by a theory and the dependent variable. Researchers are then able to determine whether these mechanisms were in fact "activated" but countervailed against, or were completely inoperative.

Third, the researcher determines how difficult a test a case poses for a specific theory by *identifying the strength of countervailing conditions present*. Say a researcher drew on a formulation of democratic peace theory to argue that disputes between democratic states should be resolved peacefully. In a dispute between two democracies, a countervailing condition against the expected outcome—peaceful resolution—might be a rapid shift in the balance of

[11] Eckstein, "Case Study and Theory," 88-90; George and Bennett, *Case Studies and Theory Development*, 116; John Gerring, *Case Study Research: Principles and Practices* (New York: Cambridge University Press, 2007), 116; Stephen Van Evera, *Guide to Methods for Students of Political Science* (Ithaca: Cornell University Press, 1997), 31.

power that made credible commitments problematic.[12] Countervailing conditions should be derived from well-specified alternative theories and supported by evidence demonstrating they have the effects the claimed for them.

According to the CC rationale, a case is LL if the researcher can identify a set of variables that strongly countervail against the outcome the theory of interest expects to occur. The more countervailing conditions the author can identify, and the stronger the logic and evidence establishing why these conditions would lead to an outcome that diverges from that expected by the theory of interest, the less likely one is to observe the outcome that theory predicts in the chosen case. LL cases have been described as "hyper-strong" and "unfair" tests because the variables highlighted by a valid theory might not appear to affect outcomes given the presence of strong countervailing conditions.[13] However, LL cases are only unfair tests if used to disconfirm theories, and should not be used for such purposes. Although failure in an LL case does not undermine a theory, success strongly indicates its generalizability: if a theory's expectations are met even when numerous conditions in a case indicate some divergent outcome should occur (e.g., war rather than peace), one would expect the theory could explain comparable cases where countervailing conditions were not as strong. Analogously, observing that someone ran and completed a marathon allows us to confidently assert that, ceteris paribus, they could also run and complete shorter races. Conversely, failure in a hard case by no means eliminates the possibility that the outcome a theory expects would occur in easier cases, just as knowledge that someone failed in their attempt to finish a marathon does not preclude the possibility they could complete a shorter distance.

---

[12] Robert Powell, *In the Shadow of Power: States and Strategies in International Politics* (Princeton: Princeton University Press, 1999); Stephen Van Evera, *Causes of War: Power and the Roots of Conflict* (Ithaca: Cornell University Press, 1999).

[13] Van Evera, *Guide to Methods*, 34.

For a case to qualify as ML, a researcher needs to identify conditions that a theory expects to exert particularly strong effects on the outcome of interest, and make sure the case features few if any countervailing conditions identified by alternative theories. Cases whose antecedent conditions lead the theory of interest to expect extreme values of the dependent variable are good candidates for selection as ML, as the effects of the theory's explanatory variables are more likely to be detected.[14] Success in an ML case does not provide strong support for a theory, but failure indicates it is unlikely to explain the outcome of interest in comparable cases. To use the running analogy, knowledge that someone could not complete a run around a mile-long track largely precludes the possibility that person could run around the same track 26 times consecutively.

INSERT FIGURE 1 HERE

This is represented in Figure 1. The y-axis represents the theory of interest a researcher is investigating, while the x-axis represents an alternative theory. Cases are arrayed within the plane formed by these axes according to how strongly each theory would expect an outcome to occur given the values of the variables in a case. Outcomes could be categorical, such as the presence or absence of a military alliance between two states, or continuous, such as the percentage change in a country's military spending in a given year. The farther a case falls above the x-axis, the more strongly the variables deemed relevant by the theory of interest predict a "positive" outcome (e.g. a large increase in military spending or high probability of an

[14] Eckstein, "Case Study and Theory," 119; George and Bennett, *Case Studies and Theory Development*, 121; Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994), 209; Jack S. Levy, "Qualitative Methods in International Relations," in *Millennial Reflections on International Studies*, eds. Michael Brecher and Frank P. Harvey (Ann Arbor: University of Michigan Press, 2002), 442.

alliance); conversely, the theory predicts increasingly "negative" outcomes the farther a case falls below the x-axis. Likewise, the alternative theory expects increasingly positive outcomes for cases falling farther to the right of the y-axis, and negative outcomes for cases to the left. The theories each yield less certain predictions for cases falling near the intersection of the two axes. Their respective variables are at values expected to have weak effects on the outcome of interest, and random measurement error is more likely to "push" the case from one quadrant of the plane to another. Because LL/ML cases require theories with specific expectations about what outcome should occur, cases near the origin in Figure 1 should be avoided.

ML cases fall in the northeast and southwest quadrants of the plane because the alternative reinforces the expectations of the theory of interest that a positive or negative outcome will occur, respectively. Since there are no countervailing pressures in these quadrants, the farther a case moves from the x-axis the more strongly the theory of interest expects a positive or negative outcome in a given case. Cases become increasingly ML, or "easy", and the inferences one can draw if the outcome the theory of interest expects does *not* occur in these cases thus become increasingly strong.

The northwest and southeast quadrants are more complicated. Here, the case variables highlighted by the alternative explanation countervail against the outcome expected by the theory of interest at varying levels of intensity. Along the dashed diagonal line, the variables highlighted by both theories are at levels that roughly offset one another. For instance, in the northwest quadrant the theory of interest predicts a positive outcome in each case along the diagonal just as strongly as the alternative predicts a negative outcome. The farther a case falls off the diagonal to the left in the northwest quadrant, the more strongly the variables highlighted by the alternative countervail against the outcome the theory of interest predicts. These cases are

increasingly LL, and the inferences one can draw if the outcome the theory of interest expects *does* occur in these cases thus become increasingly strong. Michels' cases fall in this northwestern region: his theory predicts the "positive" outcome of oligarchic governance even though other factors strongly predict democratic practices to dominate.

Conversely, the farther a case falls off the diagonal to the right in the northwest quadrant, or to the left in the southeast, the less strongly the variables highlighted by the alternative countervail against the outcome the theory of interest predicts. If a case falls very far up the y-axis—what could be called an "extreme" case—and far off the diagonal, it is an ML case for the theory of interest even though there are weakly countervailing pressures present. An extreme case is one in which the key variables in a theory adopt values one would expect to have a substantially greater-than-average impact on the dependent variable, which should then take on a value far from the norm in comparable cases.[15] George and Bennett, as well as Levy, characterize Allison's study of the Cuban Missile Crisis in the manner just described.[16] It is an extreme case for the unitary actor model of national security behavior. Government action should be much more unified than in non-crisis situations, when parochial bureaucracies would be expected to have more sway. Conversely, the crisis is not an extreme case for the organizational and bureaucratic models of national security behavior, which expect disjointed decision-making based on differences in government agencies' established routines and interests. Both theories' expectations countervail against one another. However, in contrast to the unitary actor model, there was nothing about the crisis in Cuba expected to augment the effects of standard operating

[15] Jason Seawright and John Gerring, "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options," *Political Research Quarterly* 61, no. 2 (2008): 301-302. The relationship between a theory's independent and dependent variables in an extreme case may not be monotonic; more extreme values of the independent variables need not lead to more extreme outcomes.

[16] George and Bennett, *Case Studies and Theory Development*, 122-123; Jack S. Levy, "Case Studies: Types, Designs, and Logics of Inference," *Conflict Management and Peace Science* 25, no. 1 (2008): 12-13; Graham T. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis* (Boston: Little, Brown, 1971).

procedures or bureaucratic interests on decision-making. The crisis is thus ML for the unitary actor model and LL for the alternatives Allison explores.[17] If the presence of unified decision-making is the dependent variable and the unitary model treated as the theory of interest, the Cuban Missile Crisis falls above the diagonal in the northwest quadrant of Figure 1.

Scope Conditions

Scope conditions define the set of cases to which a theory is applicable, meaning its explanatory variables are expected to be able to affect the outcome of interest.[18] Whether or not these effects actually manifest themselves as expected by the theory is an empirical question. To be meaningful, the claim that a case is LL/ML should refer to its difficulty for a theory relative to other cases also within that theory's scope conditions. This may seem obvious, but scholars have made confusing pronouncements on this point. Van Evera defines an LL test as one in which a researcher "evaluates theories under circumstances that lack the antecedent conditions [the theories] require to operate," but such cases are by definition beyond a theory's scope and thus provide no information pertinent to the theory's validity.[19] To use an absurd example, if a theory meant to explain war onset does not account for the amount of compensatory damages awarded by juries in injury litigation, this clearly does not impugn the theory. Unfortunately, violations of a theory's scope conditions are never so obvious. However, if one is testing a well-specified theory it should be possible to determine which conditions are necessary for the theory to be

[17] Even if scholars agree that the missile crisis was an LL case for the bureaucratic and organizational models of foreign policy behavior, there have been strong criticisms of Allison's conclusions based on his interpretation and use of evidence from the case. Consensus regarding the strength of countervailing conditions, and thus the difficulty of a test, need not be accompanied by consensus about how well a particular theory accounts for case outcomes. See Stephen D. Krasner, "Are Bureaucracies Important? (Or Allison Wonderland)," *Foreign Policy*, no. 7 (Summer 1972): 159-179; David A. Welch, "The Organizational Process and Bureaucratic Politics Paradigms: Retrospect and Prospect," *International Security* 17, no. 2 (1992): 112-146.

[18] James Mahoney and Gary Goertz, "The Possibility Principle: Choosing Negative Cases in Comparative Research," *American Political Science Review* 98, no. 4 (2004): 660-661.

[19] Van Evera, *Guide to Methods*, p. 34.

applicable, and also what nominal scope conditions are in fact arbitrary restrictions meant to protect the theory.

If a theory "succeeds" in explaining the outcome in a case beyond its scope conditions the researcher might have missed an important piece of evidence demonstrating the case actually falls within the theory's scope. This might also indicate the theory is poorly specified, meaning the researcher has excluded background conditions and/or causal pathways by which its explanatory variables can influence the outcome being studied.[20] If that is so, then "success" in this case actually requires revision of the theory, which can hardly be considered a confirmatory result. An under-specified theory also raises the question of whether its expectations for the case in question were precise enough to justify calling it an LL or ML test in the first place.

More ambiguously, Levy contends that an LL case is one "in which the theory's scope conditions are satisfied weakly if at all."[21] The sizeable problems introduced by the clause "if at all" have been discussed—if a case does not at all fall within a theory's scope conditions, the LL label cannot apply. It is less clear what the consequences of a theory's scope conditions being "satisfied weakly" are. Scope conditions might be weakly satisfied if there remains some uncertainty as to whether or not a case actually falls within them. If the researcher fails to realize they have chosen a case outside the theory's scope conditions, they may erroneously conclude their theoretical expectations were not met because of strong countervailing conditions rather than the theory being inapplicable. This will lead them to draw improper inferences about the relative strength of different variables' causal effects, when in actuality no meaningful information about the theory can be ascertained from the case. In the face of uncertainty

[20] Such "weak" theories make it more likely the researcher will analyze irrelevant cases or exclude relevant ones; Mahoney and Goertz, "The Possibility Principle," 661.

[21] Levy, "Qualitative Methods in International Relations," 442; Jack S. Levy, "Qualitative Methods and Cross-Method Dialogue in Political Science," *Comparative Political Studies* 40, no. 2 (2007): 202.

pertaining to scope conditions, more fine-grained theory specification and exploratory research is merited before LL/ML cases are selected. Still, a Bayesian interpretation of weakly satisfied scope conditions can more sensibly categorize a case as LL than can the CC framework, if "weakly satisfied" is taken to mean an existing theory is being applied to a new empirical domain. This is expanded upon in the section on Bayesianism below.

To elaborate upon the issue of scope conditions and LL/ML studies, return to the question of whether cases relevant to national security are particularly hard for constructivist theories to explain. The contributors to *The Culture of National Security* claim to focus on "hard cases… that favor well-established perspectives in the field of national security." As becomes clear, the alternative perspectives are structural theories, particularly neorealism.[22] Desch soon challenged the contention that constructivists had selected LL cases to test their theories: citing Eckstein, he argued constructivist scholarship had actually used "most likely" cases for "culturalist theories" and "least likely cases for the realist alternative."[23] Setting aside the inability of scholars to agree on the basic features of a case that would make it LL or ML, lost in the debate was the important question of neorealism's scope.

Neorealism continues to be one of the primary foils for constructivist theories of security, from which purported countervailing conditions are drawn. However, concurrent with emerging constructivist approaches to security studies, realists themselves were enmeshed in a debate over whether any specific case of foreign policy-making fell within the scope of structural versions of the theory. Waltz consistently maintained they did not: neorealism could only explain outcomes

[22] Katzenstein, "Alternative Perspectives on National Security," 11, 12-17.

[23] For example, Desch contends it is unsurprising that humanitarian concerns best explain policy choices in cases studied by constructivists, as the material concerns realists argue policymakers to prioritize above all others did not countervail against other values at issue; Desch, "Culture Clash," 159-160.

at the international level of analysis.[24] If so, then cases could not be categorized as LL by dint of countervailing conditions derived from neorealism, because the theory would be inapplicable to specific cases of foreign policy. Ironically, constructivist scholars who employ neorealism to justify the LL label for cases sometimes endorse this point. Declaring cases of national security policy "hard" because of neorealist theory, Katzenstein continues that "Neorealism is too general and underspecified to tell us anything about the direction of balancing, let alone the content of the national security policies of states." At best, he states, scholars wishing to apply neorealism to specific cases have to "graft" additional variables onto the theory in an ad hoc fashion, resulting in a substantially modified explanatory framework.[25] In short, Katzenstein agrees with Waltz that unless neorealism is substantially modified, specific cases of foreign policy are beyond its scope. Likewise, Kier contends that developments in French military doctrine in the 1920s and 1930s should be an "easy case" for structural realism to explain, but then contradicts this argument by noting that the international balance of power—the theory's dominant explanatory variable—"does not provide determinate explanations for choices between offensive and defensive doctrines."[26] If the balance of power never yields specific predictions of the substance of states' military doctrines, it is fair to say the question is beyond neorealism's scope. Such a claim is suspect, per the work of Posen and others.[27] But if Kier only means the status of the balance of power does not lend itself to precise predictions in the French case, then it still

---

[24] See Colin Elman, "Horses for Courses: Why Not Neorealist Theories of Foreign Policy?" Security Studies 6, no. 1 (1996): 7-53; and, in reply, Kenneth N. Waltz, "International Politics is Not Foreign Policy," Security Studies 6, no. 1 (1996): 54-57.

[25] Katzenstein, "Alternative Perspectives on National Security," 26.

[26] Elizabeth Kier, "Culture and Military Doctrine: France between the Wars," *International Security* 19, no. 4 (1995): 84. This accords with Levy's observation that implications of neorealism at the state level of analysis are hard to discern; Jack S. Levy, "Balances and Balancing: Concepts, Propositions, and Research Design," in *Realism and the Balancing of Power: A New Debate*, eds. John A. Vasquez and Colin Elman (Englewood Cliffs, NJ: Prentice-Hall, 2002), 128-153.

[27] Barry R. Posen, *The Sources of Military Doctrine: France, Britain and Germany Between the World Wars* (Ithaca: Cornell University Press, 1984).

cannot be said to strongly countervail against the effects of variables highlighted by her own theory. Both Katzenstein and Kier's points undercut the inferential leverage their otherwise valuable scholarship can claim.

Case-specific Expectations

Carefully specifying a theory and checking to see whether a case fits within its scope does not guarantee that one can derive precise expectations of what outcome should be observed if the theory is correct. This need not mean the theory itself contains logical contradictions. Instead, different variables highlighted by the theory may actually countervail against each other in a specific case. For example, Theis examines the foreign policy behavior of the United States from 1783-1814 to test whether neorealism accurately explains how states are socialized into adopting particular types of security behavior. Setting aside the issue of neorealism's scope, Theis argues that the relative weakness of the U.S. during this period should have made the country adopt "an overriding concern with the external environment and potential threats to its survival," perhaps making it an "easy case" for neorealist accounts of socialization. However, he also observes that the United States' geographic distance from great powers in Europe limited opportunities for socialization and arguably makes the early U.S. a "hard case" for neorealism.[28]

Theis's interpretation of neorealism highlights two different factors, military strength and the intensity of state interactions with great powers, that countervail against one another in the case he has chosen. This hinders the researcher's ability to say what outcome should be expected in the chosen case, and thus neither the LL nor ML label can be confidently applied. Generally, the more variables a theory expects to affect the outcome of interest, the harder it is to select

[28] Cameron G. Thies, "State Socialization and Structural Realism," *Security Studies* 19, no. 4 (2010): 703-704.

particular cases with which to conduct LL/ML tests. Complex explanations will often contain numerous countervailing conditions when applied to specific cases, making it much harder to determine what outcome the theory expects. Complex theories exempt from this claim would have to highlight variables which were **a)** positively correlated with one another and **b)** all either positively or negatively correlated with the dependent variable rather than having opposing effects. Outside these rare instances, parsimonious theories are more easily subjected to LL/ML tests.

Although cases for which a theory does not have precise expectations cannot pose easy or difficult tests of that theory, they can be used to learn about the relative impact of different variables the theory highlights. In the study just described, Theis concludes that the young United States was socialized to behave in a manner consistent with neorealism. This finding comes despite the fact that the country's distance from the great powers of the era arguably limited opportunities for the U.S to be socialized to adopt realist security practices. Thus, one might tentatively conclude that a state's relative power influences its susceptibility to socialization more than the density of its interactions with other states.[29] The implications of this finding could then be further examined in regards to related variants of realism, such as "balance of threat" theory, that stress both geography (with its implications for state interactions) and military power as crucial factors influencing alliance formation.[30]

Requiring precise theoretical expectations for a given case is not equivalent to requiring a particular outcome to occur in that case for it to count as LL/ML. If a theory predicts that outcome Y will be observed in a case, and Y does in fact occur, this has no more bearing on the case's "hardness" than if some other outcome had been observed. Rather, it is the strength of

[29] Further research would be necessary to explore this possibility. Perhaps only minimal interactions with great powers can socialize states to behave according to neorealist precepts regardless of their relative strength.

[30] Stephen M. Walt, *The Origins of Alliances* (Ithaca: Cornell University Press, 1987).

countervailing conditions that matters when defining cases. However, there is confusion in the literature on this point. Specifically, Gerring defines an LL case as "one that, on all dimensions *except* the dimension of theoretical interest, is predicted not to achieve a certain outcome, and yet does so." Gerring insists that for a case to be LL the outcome one would expect from the dimension of interest *must* occur. Similarly, he asserts for a case to be ML, the outcome one would expect from the dimensions that are *not* of interest *cannot* occur: "A 'most-likely' case is one that, on all dimensions *except* the dimension of theoretical interest, is predicted to achieve a certain outcome, and yet does not."[31] For practical purposes these definitions are logically indistinguishable. In both instances the antecedent conditions in the case countervail against the expectation yielded by the "dimension of theoretical interest." In neither definition is the outcome expected by the theory of interest consistent with that expected by the countervailing dimensions. These are both essentially definitions of LL cases. The only difference is that the latter definition allows for a much wider set of outcomes: the outcome expected by the dimensions other than that of theoretical interest cannot occur, but any other outcome is allowed. After all, it does not follow that the outcome expected by the theory of interest must be observed when other theories fail.

There are further reasons why observed outcomes should not define cases. If a theory fails to account for the outcome in an LL case researchers cannot conclude it is strongly disconfirmed, but may better infer how strongly the variables it highlights impact the dependent variable relative to those in alternative theories. If a theory—even one not primarily of interest to a researcher—passes an easy test in an ML case it may be worth testing it against a broader set of evidence to examine how generalizable it is. Furthermore, selecting ML cases in which the outcome favored by a theory does occur gives the researcher the opportunity to examine its

[31] Gerring, *Case Study Research*, 115 (emphasis in original).

hypothesized intervening causal mechanisms, which should be highly visible. Fravel employs this "modified most likely case study research design" in his research on diversionary war. He examines Argentina in 1982 and Turkey in 1974, which should be ML cases for diversionary theory due to high levels of political unrest and other facilitating conditions. The theory's expected outcome is observed in both cases: Argentina invaded the Falkland (Malvinas) islands, and Turkey invaded Cyprus. Fravel correctly notes that when a theory's key explanatory variables take on values strongly expected to lead to a specific case outcome, and that outcome is in fact observed, the intervening mechanisms hypothesized by the theory can also be expected to be present. If process-tracing nevertheless fails to uncover these mechanisms, this is strong disconfirmatory evidence for the theory in an ML case.[32]

Fravel's use of process-tracing in an ML case also illustrates a problem with the assertion that the "most likely design is intended to test *whether* X affects Y, rather than *how* X affects Y."[33] As long as the theory in question carefully specifies the intervening mechanisms between X and Y, the distinction between "whether" and "how" is one without a difference. If the expected value of the primary dependent variable is observed, the researcher can then choose as a new dependent variable the intervening mechanism between X and Y that should also be affected as the theory expects given the case outcome.[34] Having a good sense of a theory's intervening mechanisms is similarly useful in LL cases where the expected outcome is not observed. One can then determine whether these mechanisms were in fact operating but

[32] M. Taylor Fravel, "The Limits of Diversion: Rethinking Internal and External Conflict," *Security Studies* 19, no. 2 (2010): 310, 314. Marc Trachtenberg uses a similar research design in "Audience Costs: An Historical Analysis," *Security Studies* 21, no. 1 (2012): 3-42.

[33] Erik Gartzke and Yonatan Lupu, "Still Looking for Audience Costs," *Security Studies* 21, no. 3 (2012): 395 (emphasis in original).

[34] Difficulties with such a strategy arise when a theory allows for multiple different causal pathways by which the explanatory variables can influence the primary outcome of interest. Again, this highlights the importance of insuring that very few states of the world can be consistent with the expectations of the theory being tested.

countered by the effects of variables emphasized by alternative theories, or the mechanisms never even manifested themselves and thus countervailing conditions were not necessary to understand the theory's failure. The latter scenario suggests the reach of the theory is more circumscribed than does the former.[35]

Identifying Countervailing Conditions

Case studies are more adept at finding the "causes of effects," or the processes that influence an outcome of interest, than the "effects of causes", or the average effect some variable has on an outcome.[36] Nevertheless, the claim that a case's conditions strongly or weakly countervail against the expectations of the theory of interest presupposes that one can arrange the causal variables according to the strength of their effects on the dependent variable. Although researchers using cases studies will not assign numerical coefficients to variables, as with regression analysis, they do need to justify claims as to whether or not some variable will have a strong countervailing effect vis-à-vis variables highlighted by the theory of interest. This is why such claims should be based on pre-existing theories supported by empirical research that provides a reliable idea of whether different variables should have strong, moderate, or weak effects on the outcome under investigation.[37]

Although this advice is often reflected in security studies research, scholars sometimes claim a case is LL without citing any rival theory or research to support these assertions. One possible reason for the occasional neglect of research featuring alternative theories may be that

---

[35] I thank an anonymous reviewer for this point.

[36] Gary Goertz and James Mahoney, *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences* (Princeton: Princeton University Press, 2012), ch. 3.

[37] Timothy J. McKeown, "Case Studies and the Statistical Worldview," *International Organization* 53, no. 1 (1999): 182; George and Bennett, *Case Studies and Theory Development*, 121; Desch, "Culture Clash," 166; Jack Snyder, "Richness, Rigor, and Relevance in the Study of Soviet Foreign Policy," *International Security* 9, no. 3 (1984/1985): 106; Van Evera, *Guide to Methods*, 31.

prominent works on case study methods employ definitions that are easy to misinterpret. For instance, George and Bennett define an LL case as one in which "the independent variables in a theory are at values that only weakly predict an outcome or predict a low-magnitude outcome."[38] Researchers looking to follow this definition are posed with two difficulties. First, it is unclear why a case presents a difficult test simply because a theory predicts a low-magnitude outcome. Plausible interpretations of "low-magnitude" is that some outcome is either absent (e.g. "no war occurred") or well below its average value in comparable cases ("there were very few casualties during the military intervention"). If key variables specified by a theory lead the researcher to expect a low-magnitude outcome in a case, and no factors are identified which countervail against this expectation, then the case is in fact ML for that theory—it would fall within the southwest quadrant of Figure 1. Observing a "high-magnitude" outcome—large numbers of casualties in an intervention when the theory predicts few will occur, for instance—would be disconfirmatory evidence. Furthermore, a case could be LL for a theory in which it predicts a high-magnitude outcome. If a theory expects a military intervention to result in greater than average casualties, but alternative theories cite strong reasons for casualties to be limited or non-existent, this intervention would be an LL case for the first theory given the CC rationale, and it would fall below the diagonal in the northwest quadrant in Figure 1. To reiterate, outcomes alone do not define LL/ML cases.

A greater difficulty arises in trying to interpret what is meant by "weakly predict." If a theory only weakly predicts outcome Y, then Y should have a low probability of occurring. This might be because the theory is largely indeterminate for a given case—Y is just one of multiple outcomes that could be reconciled with the theory. Similarly, Levy holds that a case may be LL if the "values of many of the theory's key variables point in [the] other direction" from the value

[38] George and Bennett, *Case Studies and Theory Development*, 121.

of the dependent variable expected.[39] Again, the case then probably does not allow the researcher to say whether the theory's expectations are met, as it contains numerous variables the theory expects to countervail against one another in that instance, and the LL label cannot fit. It is unlikely this is George and Bennett's meaning, however, as they argue that for a theory to be subjected to "strong tests" it should "posit high-probability… relations between variables of interest" and forbid the possibility of numerous different outcomes in any case.[40] The case might also strongly undermine the theory being tested, a possibility that arises from Levy's definition. If numerous elements of a theory indicate the outcome of interest is unlikely to be observed in a case, and the outcome nevertheless occurs, this clearly calls the generalizability of the theory into question. This is why LL cases must derive countervailing conditions from alternatives to the main theory being tested: to avoid indeterminate expectations and erroneous inferences.

To further illustrate the confusion that can result from ambiguous definitions of LL cases, consider Kaufman's largely well-done research on ethnic violence. In one article Kaufman evaluates three theories, including an elite-predation explanation in which ethnic leaders seek to secure power by provoking intergroup conflict, and a theory of symbolic politics that holds violence is driven by myths that justify hostility towards out-groups.[41] Citing George and Bennett's definition of LL cases, he argues that the emergence of ethnic violence in Sudan in 1983 is a "tough test" for both theories "because key causal factors identified by each theory are questionable." According to Kaufman, the Sudanese leader in 1983 was unlikely to become a predatory leader or encourage hostile symbolic narratives towards outgroups.[42] However, contrary to Kaufman's assertion, this either makes the Sudanese case indeterminate or ML for

---

[39] Levy, "Qualitative Methods," 202.

[40] George and Bennett, *Case Studies and Theory Development*, 116.

[41] Stuart J. Kaufman, "Symbolic Politics or Rational Choice? Testing Theories of Extreme Ethnic Violence," *International Security* 30, no. 4 (2006): 45-86.

[42] Kaufman, "Symbolic Politics or Rational Choice," 56.

both theories. If the "questionable" status of the theories' key variables is taken to mean they are measured with great uncertainty—measurement error being another source of weak predictions—or at levels that are insufficient to have a reliable impact on levels of ethnic violence, then one cannot derive precise expectations from either theory and the LL label cannot apply. Conversely, it could be plausibly argued that both theories expect the same low-magnitude outcome—"no violence"—to occur. They thus do not countervail against each other, and the ML label would be more fitting. Because both theories predict a negative outcome, the absence of ethnic violence, the Sudan case falls in the southwest quadrant of Figure 1.

The fundamental issue with George and Bennett's definition of LL cases is that it can lead researchers to confuse the *precision* of a theory's expectations with the *strength of the causal effects* posited by the theory's key variables. A more easily understood revision of their initial wording, consistent with George and Bennett's broader methodological discussion, would eschew references to the magnitude of outcomes or the weakness of predictions. Instead, it would state that an LL case is one in which the independent variables in a theory are at values that predict a specific outcome, but not at values expected to have a stronger-than-normal effect on the dependent variable. Additionally, the case must feature conditions exogenous to the theory that would lead one to expect a different outcome. Given George and Bennett's emphasis on the importance of taking the expectations of alternative theories into account, as well as constructing theories with precise expectations, this is the clearest and most plausible interpretation of their meaning.

THE BAYESIAN RATIONALE

The CC rationale for selecting cases includes an implicit Bayesian element. The claim

that a case is an LL test for a theory because one can derive strong countervailing conditions from another theoretical explanation implies that, a priori, the researcher believes there is a reasonable probability the alternative is a valid explanation for the outcome of interest. If this were not so, the nominal "countervailing" conditions in the case would not actually pose a hard test for the researcher's theory. Eckstein hints at a Bayesian rationale for selecting cases by arguing that a case may be LL/ML based on "prevailing theory."[43] Similarly, Snyder writes that for a case to qualify as LL for a researcher's theory, the competing theory should have "home-court advantage."[44] This is to say the case belongs to a broader class for which the competing theory has been demonstrated to be a plausible explanation for the outcome of interest, while the researcher's own theory has not. McKeown argues this is an informal "folk Bayesian perspective," and he and others have more thorough in detailing this logic as applied to LL/ML cases.[45]

Bayesian inference rests on three "legs": the strength of one's prior beliefs about a theory; how closely evidence in a case fits theoretical expectations; and the typicality of the evidence in a case. In formal terms, Bayesian tests of a theory are evaluated according to

$$P(T|E) = \frac{P(T)P(E|T)}{P(E)}$$

Researchers update their beliefs about a theory given the evidence E in a case— $P(T|E)$, their "posterior" beliefs. Prior confidence in the theory—$P(T)$—is multiplied by the likelihood one would observe E in the chosen case given the expectations of the theory, or $P(E|T)$. This

[43] Eckstein, "Case Study and Theory," 119.

[44] Snyder, "Richness, Rigor, and Relevance," 106.

[45] McKeown, "Case Studies and the Statistical Worldview," 180. See also Andrew Bennett, "Process Tracing: A Bayesian Perspective," in *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier (New York: Oxford University Press, 2008), 702-721; Andrew Bennett and Colin Elman, "Case Study Methods in the International Relations Subfield," *Comparative Political Studies* 40, no. 2 (2007): 173-174; George and Bennett, *Case Studies and Theory Development*, 9, 175; Gerring, *Case Study Research*, 116; Levy, "Qualitative Methods in International Relations," 441-442; Levy, "Qualitative Methods and Cross-Method Dialogue," 202-203.

evaluation is divided by the overall probability of observing *E* at all given its prevalence in the population of interest—the broader class from which the case is drawn. *P*(*E*) captures all the ways the evidence might have been generated in addition to the mechanisms specified by the theory of interest. To see this formally, it helps to re-write Bayes' theorem as

$$P(T_i|E) = \frac{P(T_i)P(E|T_i)}{\sum_{j=1}^{n} P(T_j)P(E|T_j)}$$

The denominator shows there are *n* theories indexed by *j* that can generate the evidence *E* with varying probabilities, including the theory of interest T_i. Summed together, these probabilities equal *P*(*E*). The values of the posterior and numerator change as the researcher goes back and forth between theory and evidence in different cases, while the value of the denominator is constant.

All else being equal, posterior confidence in a theory is greater if prior confidence was high or if the "fit" between theory and evidence in the chosen case is strong—especially if there is fit despite countervailing conditions. One's confidence in the theory also increases if the evidence one observes rarely occurs in comparable cases. Conversely, it is not impressive if one's theory is only supported by commonplace occurrences that one would expect by chance alone. For instance, one observable implication of the argument that states are engaging in "soft balancing" against the U.S. is that countries will take diplomatic actions to impede the realization of American foreign policy goals.[46] Skeptics of this argument have countered that such actions are ubiquitous: they are part of "unipolar politics as usual" or "normal diplomatic friction."[47] In short, this evidence cited in favor of soft balancing is so commonplace that we are likely to observe it regardless of whether soft balancing is taking place or not. Those who doubt the

[46] Robert Pape, "Soft Balancing against the United States," *International Security* 30, no. 1 (2005): 7-45.

[47] Keir A. Lieber and Gerard Alexander, "Waiting for Balancing: Why the World Is Not Pushing Back," *International Security* 30, no. 1 (2005): 109; Stephen G. Brooks and William C. Wohlforth, "Hard Times for Soft Balancing," *International Security* 30, no. 1 (2005): 79.

validity of the soft-balancing proposition would presumably be more persuaded if it accounted for unusual international events that one would not expect due to chance. Even unusual evidence that only seems explicable in light of a given theory must also be able to be reliably interpreted, accurate, and provide information about alternative explanations that could have generated the observed outcome. If the methods used to gather and interpret the available evidence are suspect, one should not substantially update their posterior confidence in a theory.[48] As with the CC approach to LL/ML cases, the possibility of measurement error means the degree to which countervailing pressures are present or absent, as well arguments about whether a theory's expectations are met in a case, might both be subject to substantial revision.

There are five general steps researchers using the Bayesian framework must go through when selecting cases. The three steps used in the CC rationale—checking scope conditions, making sure theories have clear and precise expectations for the case at hand, and identifying the strength of countervailing conditions in the case—all still apply. Fourth, researchers must evaluate the strength of their prior confidence in all theories being examined given the selected case. An ML case is one in which prior confidence in a theory is high, indicating its expectations are "highly likely to be confirmed" when tested against the case.[49] The most straight-forward way to justify confidence that a theory's expectations will be met is to demonstrate that the case strongly resembles "typical" cases in which the theory's expectations have previously been met.[50] This means it features numerous relevant conditions that were also observed in previously studied cases in which the theory was shown to be a good explanation for the outcome of interest. Whether or not the theory or competing alternatives best explain the outcome in the

[48] John Gerring, "Is There a (Viable) Crucial-Case Method?" *Comparative Political Studies* 40, no. 3 (2007): 235.

[49] Levy, "Qualitative Methods and Cross-Method Dialogue," 202.

[50] Typical cases are described in Seawright and Gerring, "Case Selection Techniques," 299-300.

newly selected case is an empirical question. Conversely, McKeown states that a "'hard case' for a theory… would be one where the prior probability of a theory being a correct explanation is low but the degree of confidence in that prior assessment is not high."[51] This translates to the assertion that a researcher's prior beliefs about the theory being tested given the chosen case are "uninformative." Any effects a theory's explanatory variables might be observed to have on the dependent variable are equally likely a priori, including the possibility that they have no impact. A researcher's prior beliefs might be uninformative if the theory is novel and untested—there are few if any prior observations with which to evaluate it. It might also be that the theory has been tested in a context that only remotely resembles that in which the researcher is interested, leading to questions about its external validity.

Lastly, after accounting for alternative theories and their prior probability, the researcher may find it necessary to re-adjust their initial prior confidence in the theory of interest. The greater the prior probability of the alternatives, the less likely the theory of interest is to prove a plausible explanation. As the Bayesian rationale is employed informally, the researcher need not put their prior beliefs in quantitative terms, but should clearly explain why one should be relatively (un)confident in a theory prior to examining the evidence in the case they have chosen to study.

Comparing the Two Rationales

Bennett and Elman explicitly state that LL cases follow a Bayesian logic, saying that outcomes must be evaluated against "extant theory" and that cases become harder tests for a theory of interest when alternative explanations for the evidence in question are "predominant." However, they complement this by noting that the hardest case for a theory is one in which "the

[51] McKeown, "Case Studies and the Statistical Worldview," 180.

variables of that theory are not at extreme levels that strongly push toward that outcome, and… all of the alternative hypotheses predict a different outcome in that case."[52] That most closely resembles the CC rationale described here, which is effectively "two-legged" Bayesianism: it takes the strength of case evidence and its closeness of fit with theoretical expectations into account, but does not consider a priori confidence in the theory of interest based on previous tests. The primary work is being done within the $P(E|T)$ term: how likely would one be to observe this case outcome, in the face of countervailing pressures (or the absence thereof), if the theory of interest were correct? Alternatively, under the Bayesian rationale the hardness of a case also depends on the existing confidence in the theory given previous tests. The prior term $P(T)$ omits the evidence E from the case the researcher has selected, and is instead based on the results of case studies or other methods of research in which the theory has already been evaluated using comparable cases.[53]

One important implication of the preceding observation is that, for new and untested theories, a case that appears ML using the CC rationale will be somewhat less likely once a Bayesian perspective is introduced. Consider a case with characteristics that should produce a specific outcome according to a researcher's theory of interest. The researcher cannot identify conditions that countervail against the key explanatory variables in the theory, which happen to be at values that should strongly affect the case outcome if the theory is right. Thus, the case is ML according to the CC rationale. However, if the theory itself is untested researchers have no reason to have prior confidence in its validity. They might find that the variables the theory purports have the most influence on the dependent variable are actually inconsequential. Maybe

[52] Bennett and Elman, "Case Study Methods in the International Relations Subfield," *Comparative Political Studies* 40, no. 2 (2007): 173.

[53] Note that E and $P(E)$ are not equivalent. The former is the substantive evidence observed in the new case, while the latter is the probability of observing that evidence in the population of interest.

it identifies important variables but does not postulate the right relationships between them. Though the case-specific evidence creates favorable conditions for the theory, this is tempered by the uninformative nature of the researcher's prior beliefs.

It might also be that the theory in question was developed and tested in a context that does not resemble the selected case, providing reason to be skeptical about its external validity. This rationale is sometimes used to justify the LL label for cases in security studies in which psychological theories of decision-making are applied. Prospect theory, for example, has been rigorously tested in numerous controlled experiments, and has been used to explain political actors' propensity towards risky policy choices in many different cases relevant to international security.[54] Nevertheless, neither psychology laboratories nor the typical participants in experiments are very representative of decision-making in international politics. Bearing this in mind, a recent study of Americans' attitudes towards the use of force claims that by testing prospect theory's propositions about individual's risk preferences "in a real-world setting, the study's methodological research design constitutes a hard case" for the explanation.[55] In other words, the researcher is skeptical of prospect theory's external validity, and his prior confidence in its ability to explain the new data is relatively weak. Levy's definition of LL cases as those in which a theory's scope conditions "are satisfied weakly" makes more sense when interpreted through this lens. There may be no compelling reasons to suspect a case falls outside a theory's scope conditions, but also little if information indicating how the theory will perform in the new empirical domain.

---

[54] Good overviews include Jack S. Levy, "Loss Aversion, Framing Effects, and International Conflict: Perspectives from Prospect Theory," in *Handbook of War Studies II*, ed. Manus I. Midlarsky (Ann Arbor: University of Michigan Press, 2000), 193-221; Jonathan Mercer, "Prospect Theory and Political Science," *Annual Review of Political Science* 8 (2005): 1-21.

[55] Héctor Perla Jr., "Explaining Public Support for the Use of Military Force: The Impact of Reference Point Framing and Prospective Decision Making," *International Organization* 65, no. 1 (2011): 152.

While it is important to keep the Bayesian perspective in mind, an LL case must feature some plausible countervailing conditions derived from alternative theories for it to provide strong inferential leverage. It is not very persuasive to contend that one should place confidence in a theory solely because it succeeds in an instance in which the researcher had uninformative priors. If that were so, then any novel theory that passed exploratory tests would be considered more likely valid than not. Instead, alternative theories should identify numerous countervailing factors, as in the CC rationale, in combination with the researcher's uninformative priors about the theory of interest. Even then, as discussed above and in greater detail in the penultimate section of the article, a case may not increase posterior confidence in a theory by much if the outcome of interest is very common.

The Bayesian perspective also qualifies the CC rationale in terms of how the number of competing alternative theories impacts the difficulty of a case. Suppose that numerous distinct theories all expect the same outcome in a case. George and Bennett argue that such a case is ML for all theories, because rather than countervailing against one another the conditions deemed relevant by each theory reinforce each other's expectations.[56] They are employing the CC rationale, and state that this is the best scenario for strong disconfirmation of the theory of interest. Conversely, from the Bayesian perspective there may be a substantial problem with this disconfirmation strategy, as the strength of alternative explanations can undermine confidence in the theory of interest even if they all expect the same outcome in an "over-determined" case. To understand why, recall that the denominator in Bayes' theorem, which is a probability expressing all the ways the observed evidence is generated in the population of interest, is a constant. If it is to remain constant, prior confidence in at least one of the other potential explanations already under consideration must decline when another plausible

[56] George and Bennett, *Case Studies and Theory Development*, 122.

alternative is identified—leading to the fifth and final step in the Bayesian process of selecting LL/ML cases.[57] Analogously, a doctor can be more confident of correctly diagnosing a patient's illness the first time if there are very few diseases that can cause the patient's symptoms, rather than very many. Researchers must use their judgment to determine whether the prior probabilities of some theories are more influenced by the introduction of new explanations than others. The greater the researcher's prior confidence in the set of alternatives, the less initial confidence in the theory of interest.

To further illustrate, consider Monten and Bennett's study of crisis decision-making during the Gulf War. The authors examine three theoretically unique models of decision-making that are "competing and at times incompatible" with one another: one centered on presidential leadership, another on bureaucratic politics, and a third on organizational culture. They then claim that the Gulf War is a "most likely case" for each of these models, as the explanatory variables cited by each exhibited extreme values in the case and thus should all be expected to strongly affect political actors' decisions.[58] There are two problems with this assertion. First, even if the variables highlighted by the three theories are all at extreme values, by the authors own assertion the theories from which they are derived mainly contradict one another, expecting different types of decision-making (the dependent variable) to predominate. Thus, the key

[57] This may be further clarified by Ingo Rohlfing, "Comparative Hypothesis Testing via Process Tracing," *Sociological Methods and Research* (2013; advance access). Rohlfing invokes the concepts of "theoretical uniqueness" and "empirical uniqueness." The former means that the causal mechanism specified by a theory is not shared by alternative theories. The latter means that the causal mechanisms specified by alternative theories cannot be in operation simultaneously in any particular case. As far as Bayesian inference is concerned, theoretical rather than empirical uniqueness is key when assessing the probability the evidence in the denominator was generated by a given theory. This is because $P(E)$ refers to the probability of observing the outcome being researched within the population of interest as a whole, not within any particular case. Thus, even if multiple theories can all affect the outcome of interest in a chosen case (i.e. they lack empirical uniqueness), their prior probability must decline as long as the mechanisms they argue produce the outcome are distinct.

[58] Jonathan Monten and Andrew Bennett, "Models of Crisis Decision Making and the 1990–91 Gulf War," *Security Studies* 19, no. 3 (2010): 487-488.

conditions argued to matter in the case countervail against each other's expected effects, making it difficult to see how it could be an ML test for any of the theories from the CC perspective.

Second, even if one could find a case in which the different models all expected the same outcome, it would still be difficult from the Bayesian perspective to place substantial prior confidence in any of them. As Monten and Bennett assert, they are theoretically incompatible with one another, and all have been advanced as equally plausible explanations for the case. There are three "suspect diseases" here, not one or two. If another plausible explanation was added to the mix, the collective prior probability of the original three theories would have to decline even further. Helpfully, Monten and Bennett further disaggregate the Gulf War into multiple cases, or "phases", and argue that certain theories are more likely to explain decision-making in some phases rather than others. This is a useful technique to apply when theories appear equally plausible explanations for the broader case, as in the example of the Gulf War. Without disaggregation, Monten and Bennett's case would be closer to a disciplined-configurative study in which researchers apply established theories to see which better accounts for a specific outcome, not to establish the general (in)validity of the theories themselves.[59]

Bayesianism and "Surprising" Results

Levy writes "The higher the a priori probability that [a] case will satisfy the theory [in question], the greater the inferential leverage if the case does not fit the theory."[60] While this is correct within the narrower CC framework, in which prior expectations are based only on the evidence in a specific case, it is inaccurate if it refers to prior beliefs about the theory's validity for a broader class of similar cases, as in the Bayesian rationale. Relatedly, Bennett and Elman

[59] Sidney Verba, "Some Dilemmas in Comparative Research," *World Politics* 20, no. 1 (1967): 111-127.

[60] Jack S. Levy, "Coercive Threats, Audience Costs, and Case Studies," *Security Studies* 21, no. 3 (2012): 388.

argue that within a "Bayesian logic" it is true that "the more surprising an outcome is relative to extant theories, the more we increase our confidence in the theory or theories that are consistent with that outcome."[61] This argument is often correct, but the emphasis placed on fit between expectations and evidence masks considerable ambiguity concerning the implications a "surprising" outcome has for one's prior beliefs about a theory and its alternatives, as well as for the strength of evidence in the case.

One reason an outcome may be surprising is that the researcher had great prior confidence in a theory due to previous empirical investigations, and yet the theory's expectations were not met in the chosen case. This is the situation Levy describes above, if one interprets the statement from a Bayesian perspective. McKeown also claims that if a well-accepted theory does not account for the outcome in a particular case despite expectations, this finding should be treated as highly instructive.[62] All else being equal, larger errors in prediction do lead to larger changes in confidence. However, if a negative finding is also very surprising because it departs from previous evidence supporting a theory, this indicates the researcher had considerable confidence in the theory before examining the new case. The unexpected results are going to be offset by the researcher's strong prior beliefs, and the stronger the prior evidence, the less the researcher should alter their confidence in the theory even if its prediction was wrong.[63] For example, some research supports a dynamic interpretation of the democratic peace in which democracies gradually learn they can peacefully reconcile their disagreements.[64] An observable implication is that democratic dyads should be more likely to reconcile their conflicts peacefully

---

[61] Bennett and Elman, "Case Study Methods," 173.

[62] McKeown, "Case Studies and the Statistical Worldview," 181.

[63] For a technical explanation see Simon Jackman, "Bayesian Analysis for Political Research," *Annual Review of Political Science* 7 (2004): 488-491.

[64] Lars-Erik Cederman, "Back to Kant: Reinterpreting the Democratic Peace as a Macrohistorical Learning Process," *American Political Science Review* 95, no. 1: 15-31.

over time. If researchers do observe that peaceful reconciliations of conflicts between democracies gradually become predominant, their confidence in the theory should grow. A militarized dispute or war between two mature democratic countries would then be surprising. Within the CC rationale, one might conclude the theory failed in an ML case. However, the case would be a surprise not only because it upset the theory's prediction, but also because it constituted an exception to the bulk of evidence already gathered. From the Bayesian perspective, the interesting divergent observation would nevertheless do little to undermine confidence in the theory as a whole.

If anything, results that are surprising in light of existing studies should first lead the researcher to check for research errors and re-evaluate the strength of the evidence in the case—does it appear comprehensive, accurate, and open to reliable interpretation? Even if the answer to all these questions is "yes," only additional findings contrary to an established theory's expectations will suffice to substantially weaken confidence in it. However, such cases are still useful as "deviant" observations that may eventually generate new explanations for an outcome of interest, and/or better define the scope conditions of the established theory in question. [65]

The results of a case may also be surprising if the theory's expectations were met, and the researcher had very low prior confidence in the theory in question. Confidence may be low because the theory has repeatedly failed to account for the evidence in comparable cases, meaning the researcher has informative priors about the theory's *invalidity* for that class of cases. Here, the greater the surprise derived from the results in a case, the stronger the prior beliefs in the theory's invalidity. Just as it would be erroneous to dismiss an established theory based on one failure, one should not embrace a largely discredited theory based on a single apparent

[65] Arend Lijphart, "Comparative Politics and the Comparative Method," *American Political Science Review* 65, no. 3 (1971): 692.

success. Thus, surprising results do not necessarily lead to a meaningful increase in confidence in the theory whose expectations are met.

The preceding drives home another substantial implication of adding Bayesian insights to the CC rationale for case selection. From the Bayesian perspective researchers must be careful not to select a case that is *too hard or easy* for the theory of interest, such that their prior beliefs are so strong they cannot reasonably be impacted by the results of a single case. Even if conditions in a case strongly countervail against the theory of interest, meaning the case is LL within the CC framework, a good fit between theoretical expectations and the case outcome will not be that informative if prior confidence in the theory approaches zero. More difficult tests are not necessarily more diagnostic tests within the Bayesian framework. Likewise, if prior research merits a very strong belief that a theory is valid for a population of cases, no single one of those ML cases will be useful for disconfirming the theory. Although anomalous findings can be useful for theory generation, this is not the purpose of LL and ML cases, which are instead meant to provide substantial inferential leverage from a small number of observations.

As McKeown's definition of LL cases acknowledges, it is when one's priors about the theory of interest are relatively uninformative that new evidence has the largest impact on posterior beliefs.[66] Take the situation when a case outcome is surprising from the perspective of a previously established theory, but matches the expectations of a less established theory for which the researcher has uninformative priors. This is likely the scenario Bennett and Elman have in mind when discussing "surprising" findings.[67] In this scenario an unexpected result can substantially boost confidence in the unestablished theory, although the established alternative is not disconfirmed more generally. As long as prior confidence in the alternative is not so great

[66] McKeown, "Case Studies and the Statistical Worldview," 180.
[67] Bennett and Elman, "Case Study Methods," 173.

that confidence in the theory of interest was near zero entering the study, the results will be informative. Similarly, examinations of a theory within a new domain that still falls within its scope conditions may yield much stronger inferences about its validity. For instance, Wohlforth, Little, and Kaufman observe that "systemic balance-of-power theory" has been thoroughly examined in the modern European context, but not in non-European, premodern settings. They claim that studies of the latter contexts are necessary to test the theory's "transhistorical" propositions.[68] From a Bayesian perspective one could add that while new case studies of the European system may do little to alter prior beliefs about how the balance of power functions, the paucity of investigations dealing with premodern non-European systems means that these studies should have a larger impact on beliefs given researchers' uninformative priors within these domains.

NUCLEAR WEAPONS AND VIETNAM: HOW HARD A CASE?

To demonstrate how to put the preceding advice to use, this article examines Nina Tannenwald's argument that a "nuclear taboo" helps explain the non-use of nuclear weapons after 1945. A taboo is a strongly prohibitive norm against behavior socially defined as dangerous and repugnant, its injunctions taken for granted as absolute.[69] Tannenwald argues that of all relevant cases, "the Vietnam War provides one of the strongest 'tests' [i.e. an LL case] of a nuclear taboo."[70]

---

[68] William C. Wohlforth, Stuart J. Kaufman, and Richard Little, "Introduction: Balance and Hierarchy in International Systems," in *The Balance of Power in World History*, Stuart J. Kaufman, Richard Little, and William C. Wohlforth, eds. (New York: Palgrave, 2007), 5.

[69] Nina Tannenwald, *The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945* (Cambridge: Cambridge University Press, 2007), 10-11.

[70] Tannenwald, *The Nuclear Taboo*, 190. Earlier, Tannenwald called Vietnam a "hard test" for the taboo; Nina Tannenwald, "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use," *International Organization* 53, no. 3 (1999): 451.

The purpose of this section is not to question the validity of claims that a nuclear taboo existed in the minds of U.S. leaders during the Vietnam War. Rather, it is to demonstrate how to evaluate the Vietnam case to determine the degree to which it is LL. One might ask whether such an exercise is even appropriate for Tannenwald's study. Responding to the possibility that there are no true "hard tests" of the taboo's causal effects, she argues that the more interesting question is how the taboo came to be constituted. Tannenwald and colleagues have also expressed skepticism that taboos and other factors impinging on the use of particular weapons can be treated as "independent variables" that account for different proportions of variance in outcomes.[71] Nevertheless, these points are belied by Tannenwald's discussion of how to determine the taboo's "explanatory weight" and claims that "the taboo explanation [for nuclear non-use] is ultimately more powerful [than alternatives] because it captures more of the evidence."[72] Elsewhere it appears she conceives of the taboo as an interactive variable with other conditions that "would not have been as potent" in its absence, or as a "necessary condition" for outcomes such as non-use.[73] Under either conceptualization, the causal strength of the taboo can be evaluated relative to other factors affecting the outcome interest.[74] The following analysis shows that, if the steps offered in this article are followed in regards to Tannenwald's study, her contention that Vietnam is an LL case for the taboo is very much weakened. Whereas it is hard to categorize at all during the Nixon presidency, during the years Lyndon Johnson was president the relevant theories and evidence indicate Vietnam is actually an *ML* case.

---

[71] Tannenwald, *The Nuclear Taboo*, 53-54, 368; Richard Price and Nina Tannenwald, "Norms and Deterrence: The Nuclear and Chemical Weapons Taboos," in *The Culture of National Security*, 145-148. On the difference between causal and constitutive arguments, see Alexander Wendt, "On Constitution and Causation in International Relations," *Review of International Studies* 24, no. 5 (1998): 101-118.

[72] Tannenwald, *The Nuclear Taboo*, 54.

[73] Tannenwald, *The Nuclear Taboo*, 44; Price and Tannenwald, "Norms and Deterrence," 148.

[74] Although correlational analyses rather than case studies are best suited to evaluating interactive variables, the latter are useful for assessing the relative importance of necessary conditions; see Gary Goertz and Jack S. Levy, "Causal Explanation, Necessary Conditions, and Case Studies," in *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, eds. Gary Goertz and Jack S. Levy (New York: Routledge, 2007), 9-46.

Vietnam and Scope Conditions

The first step in determining whether a case can be categorized as LL or ML for a theory is to assess whether it falls within the theory's scope. Tannenwald observes that the U.S. had numerous operational nuclear weapons it could use against targets in North Vietnam if policymakers chose to do so.[75] This satisfies a minimal criterion for placing the case within her theory's scope, but the main complication involves establishing when the taboo became prevalent enough that it could have causal effects on behavior at the international level.

Norms are socially constructed over time in a non-linear fashion. It takes time for large numbers of people to become aware they exist, let alone internalize them as an obvious standard for behavior.[76] This is so for the nuclear taboo as well: Tannewald argues that it "evolved gradually 'bottom up' as a result of societal pressure," eventually affecting the beliefs and decisions of top policymakers.[77] Unless a sufficient amount of time has passed for leaders to become aware of the existence of a taboo, it cannot affect their decisions regarding nuclear weapons. Cases that fall outside this time frame will be outside the theory's scope, and rather than being LL tests will provide no useful information for assessing the theory.

It is not essential that Tannenwald specify exactly when the nuclear taboo became prevalent enough for cases to fall within its scope. Instead, she must simply show that Lyndon Johnson and his advisers were aware of it. She does so effectively, tracing the taboo's gradual emergence after the U.S. used nuclear bombs against Japan to demonstrate that by 1964 it had

---

[75] Tannenwald, *The Nuclear Taboo*, 191.

[76] Martha Finnemore and Kathryn Sikkink, "International Norm Dynamics and Political Change," *International Organization* 52, no. 4 (1998): 887-917.

[77] Tannenwald, *The Nuclear Taboo*, 56.

been "internalized" by important civilian leaders.[78] A brief allusion to the salience of the taboo in 1964, when U.S. military involvement in Vietnam began to escalate sharply, would alleviate any confusion about whether the war fell within the theory's scope conditions.

Expectations of the Taboo Theory

For a case to be an LL test of a theory, the theory must precisely specify what outcome would be consistent with its expectations. This will be easier to determine for parsimonious theories, as there will be less of a chance that numerous variables highlighted by the theory will countervail against each other in a chosen case. Though the process by which the nuclear taboo was generated was complex, its expected effect on the outcome in any case within the theory's scope is clear: policymakers should not use nuclear weapons during military conflicts. Tannenwald further details the intervening mechanisms by which the existence of the taboo should be connected to non-use. The taboo should have regulative effects that constrain behavior, constitutive effects that shape how state leaders define their interests, and permissive effects that shield non-nuclear weapons from similar levels of moral condemnation, thus promoting their use over nuclear options. These effects should be transmitted through the mechanisms of domestic and international opinion, and the moral beliefs of political leaders.[79] Not only does Tannenwald's theory have clear expectations regarding the use of nuclear weapons, then, but her careful specification of intervening mechanisms allows her to determine whether most or all were functioning in a case, or instead appeared inoperative regardless of the presence of countervailing pressures.

[78] Tannenwald, *The Nuclear Taboo*, 69.
[79] Tannenwald, *The Nuclear Taboo*, 44-51.

One way the expectations of Tannenwald's theory could be made more precise is by better articulating the relative causal weight given to intrinsic and extrinsic motivating factors. Domestic and international opinion are extrinsic: political leaders need not personally subscribe to the injunctions of a taboo to recognize that they will face costs if they violate others' sense of decency. Conversely, if leaders have internalized the taboo into their own belief systems, they will be intrinsically motivated to follow it. Tannenwald and others note that President Johnson and his advisers had strong moral qualms about the use of nuclear weapons in addition to being aware of outside opinions, while Nixon and his advisers were sensitive to opinion but lacked similar moral compunctions.[80] It is intuitive that the taboo theory would expect Nixon to have been less committed to nuclear non-use than Johnson, but to what extent is unclear. At a bare minimum, it is best to disaggregate the Vietnam War into two cases, and acknowledge that, ceteris paribus, the start of the Nixon presidency marks the beginning of a harder case for the taboo theory.

Conditions Countervailing Against the Taboo

The weakest part of Tannenwald's categorization of Vietnam as an LL case is how she identifies countervailing conditions that, contrary to the taboo theory, would lead one to expect U.S. policymakers to use nuclear weapons during the war. "In Vietnam," she writes, "the United States chose to lose a humiliating and destructive war against a small, non-nuclear adversary… [and] sustained large losses in men, money, and materiel at tremendous political cost. U.S. officials repeatedly declared that the United States could not tolerate the loss of Southeast Asia

[80] Tannenwald, *The Nuclear Taboo*, 227; T.V. Paul, *The Tradition of Non-Use of Nuclear Weapons* (Stanford: Stanford University Press, 2009), 72, 78.

to Communism, and that the war was vital for U.S. interests, prestige, and security."[81] Aside from citing the subjective stakes of the war, two assumptions are made here: first, that nuclear weapons would have been militarily useful; and second, that massive "sunk costs" would motivate policymakers to take costlier and riskier actions to secure victory. Although both arguments appear plausible, Tannenwald makes little effort to connect them to alternative theories supported by evidence. In fact, her own analysis undermines the assumptions about nuclear weapons' utility. Although members of the Joint Chiefs of Staff thought nuclear weapons might be militarily effective, other analyses were quite skeptical. The 1966 report issued by scientists in the JASON group held that tactical nuclear weapons would not be that useful in Vietnam. Though the effect of the report on senior decision-makers was "vague" according to Tannenwald, it weakens her point about the military utility of nuclear weapons in Vietnam.[82]

More problematic is the assumption about sunk costs driving escalation. This proposition is consistent with prospect theory, which would expect U.S. leaders who perceived themselves to be in a "domain of loss" given the costs of fighting would also be more disposed to take risky action to salvage victory. Some writing prior to the publication of Tannenwald's book had suggested the applicability of prospect theory to the Vietnam War and "peripheral" military conflicts fought by great powers more generally.[83] However, the bulk of extant theory and evidence contrarily indicated that the U.S. should have been unlikely to escalate based on sunk costs. "Rationalist" literature predicts that an attacker will believe the probability of victory is decreasing during protracted wars, as a weak opponent would have accepted an early settlement deal. Leaders who learn their opponent is stronger or more resolved than anticipated will

---

[81] Tannenwald, *The Nuclear Taboo*, 190.

[82] Tannenwald, *The Nuclear Taboo*, 214-218.

[83] Jeffrey W. Taliaferro, "Power Politics and the Balance of Risk: Hypotheses on Great Power Intervention in the Periphery," *Political Psychology* 25, no. 2 (2004): 202, 204; Taliaferro, *Balancing Risks: Great Power Intervention in the Periphery* (Ithaca: Cornell University Press, 2004).

decrease their war aims rather than taking risky escalatory actions.[84] Still, leaders may rationally take costly and risky escalatory actions, even if they believe the war is going badly and such actions will be suboptimal for their state as a whole, if these "gambles" increase their chances of retaining political power.[85] However, leaders of democracies like the U.S. are quite unlikely to engage in this type of behavior, presumably because they do not usually face severe punishment (prison, execution, etc.) for losing a war unless their state is conquered by the enemy. After a year-and-a-half of fighting, democracies are prone to settle for draws or losses rather than escalate.[86]

In sum, the conditions Tannenwald cites as countervailing against her own theory's expected outcome are much weaker than she portrays them given the available theory and evidence at the time. One can even reasonably place the period of the war in which Johnson was president in the ML category of cases for the taboo. Johnson was intrinsically and extrinsically motivated not to use nuclear weapons, and even without the taboo there was only a small chance that nuclear weapons would have been effective on the battlefield or protected his presidency. They were not necessary to prevent harsh punishment once Johnson left office, as U.S. leaders never feared their country was in danger of being conquered by North Vietnam during the war. Since both Tannenewald's theory and the most plausible alternatives countervail against the use of nuclear weapons during this period, Johnson's presidency falls somewhere in the southwest

---

[84] Among numerous examples, see Darren Filson and Suzanne Werner, "A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and Outcome of War," *American Journal of Political Science* 46, no. 4 (2002): 819-837; Robert Powell, "Bargaining and Learning While Fighting," *American Journal of Political Science* 48, no. 2 (2004): 344-361; R. Harrison Wagner, "Bargaining and War," *American Journal of Political Science* 44, no. 3 (2000): 469-484.

[85] Bruce Bueno de Mesquita, Randolph M. Siverson, and Gary Woller, "War and the Fate of Regimes: A Comparative Analysis," *American Political Science Review* 86, no. 3 (1992): 638-646; George W. Downs and David M. Rocke, "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War," *American Journal of Political Science* 38, no. 2 (1994): 362-380.

[86] H.E. Goemans, *War and Punishment: The Causes of War Termination and the First World War* (Princeton: Princeton University Press, 2000); D. Scott Bennett and Allan C. Stam, "The Declining Advantages of Democracy: A Combined Model of War Outcomes and Duration," *Journal of Conflict Resolution* 42, no. 3 (1998): 344-366.

quadrant of Figure 1. This does not mean that Tannenwald is wrong to argue that the nuclear taboo affected decision-making during the war, but rather one should be hesitant to infer the generalizability of the taboo's effects based on this single case.

Adding Bayesian Elements

The main contribution of the Bayesian framework to the preceding analysis is to demonstrate that, if confidence in the nuclear taboo theory is to be meaningfully increased, treating nuclear non-use as the sole outcome of interest is not going to be very useful. This is due to the (thankful) ubiquity of non-use in the relevant population of cases to which the Vietnam War is comparable. One could define the "relevant" population to include any case of militarized conflict involving at least one nuclear power. However, a more conservative approach would solely include cases of "inter-state war" in which only one side had an operational nuclear arsenal, as these are most comparable to the cases Tannenwald includes in her own studies.[87] Based on the Correlates of War dataset, this yields 14 cases, including the four Tannenwald examines.[88] In this set of cases, non-use occurs about 93 percent of the time, World War II being the only exception.

Including the nuclear taboo, Tannenwald lists five explanations derived from the disciplines of International Relations and history that plausibly account for the non-use of

[87] Tannenwald, *The Nuclear Taboo*, 69-71.

[88] Meredith Reid Sarkees and Frank Wayman, *Resort to War: 1816 – 2007* (Washington, DC: CQ Press, 2010). In chronological order, these are World War II; the Korean War; the Sinai War; the USSR's invasion of Hungary; the 1973 "Yom Kippur War"; the Sino-Vietnamese wars of 1979 and 1987; the Falklands War; Israel's 1982 invasion of Lebanon; the Gulf War; NATO's air war in Kosovo; and the U.S. invasions of Afghanistan and Iraq. A less restrictive approach would include "extra-state" wars between states and non-state entities, such as the Soviet Union's war against the Muhajadin in Afghanistan.

nuclear weapons within the relevant population of cases.[89] Given the small number of cases, and the limited number of studies on the causes of nuclear non-use at the time of Tannewald's study, it is hard to say a priori whether any explanation should be seen as more likely valid than not.[90] That is, prior beliefs about the different explanations should be largely uninformative. This is not sufficient to push any case into the LL category for the taboo theory, however, unless there is also strong countervailing conditions therein. As has been shown, theory and evidence indicates countervailing pressures were weak in Vietnam, particularly during the Johnson presidency. The typicality of evidence under examination is also crucial. Assume that prior confidence in the taboo theory, $P(T_{taboo})$, is 0.20, and for simplicity, round the probability of nuclear non-use, or $P(E_{non\text{-}use})$, to 0.90. Posterior confidence in the taboo theory after reviewing the Vietnam case would only increase if it fit the evidence very well, such that $P(E_{non\text{-}use}|T_{taboo})$ was greater than 0.90, and even then the gains in confidence would be slight.[91]

This is another reason why it is important to carefully specify intervening mechanisms in a theory. If the primary outcome of interest is overwhelmingly common in the relevant population, one might select as a new outcome one of the intervening causal mechanisms specified by the theory being tested, one that is less typical. Usually, researchers using case studies will not consider the probability of a single piece of evidence being found given the expectations of a theory, but rather the *joint* probability of observing the outcome the theory

---

[89] These exclude deterrent effects that exist when both sides in a war have nuclear weapons. Alternatives to the taboo are the uncertain long-term effects of using nuclear weapons; nuclear weapons' limited military utility; readiness factors; and the obsolescence of war. Tannenwald, *The Nuclear Taboo*, 30.

[90] Academic studies preceding Tannenwald's book include T.V. Paul, "Nuclear Taboo and War Initiation in Regional Conflicts," *Journal of Conflict Resolution* 39, no. 4 (1995): 696-717; Scott D. Sagan, "Realist Perspectives on Ethical Norms and Weapons of Mass Destruction," in *Ethics and Weapons of Mass Destruction*, eds. Sohail Hashmi and Steven Lee (New York: Cambridge University Press, 2004), 73-95.

[91] Say that the taboo theory fit the Vietnam War evidence exceptionally well despite numerous strong countervailing conditions, such that $P(E_{non\text{-}use}|T_{taboo}) = 0.99$. In this instance, $P(T_{taboo}|E_{non\text{-}use}) = P(T_{taboo})P(E_{non\text{-}use}|T_{taboo}) / P(E_{non\text{-}use}) = (0.2*0.99)/0.90 = 0.22$. Thus, posterior confidence is proportionally greater than prior confidence by ten percent. If $P(E_{non\text{-}use}|T_{taboo}) = 0.95$, posterior confidence would improve by a little less than 5.6 percent.

predicts along with the intervening mechanisms the theory proposes should produce that outcome. As joint probabilities will be less than the unique probabilities of their constitutive variables, this will yield a more decisive test.[92]

In light of all the preceding discussion in this section, a better LL case for the taboo theory would be one in which a nuclear-armed state facing a non-nuclear opponent was either a moderately repressive "mixed" regime (an "anocracy"), or a democracy that could plausibly be overrun by its adversary.[93] Given the lack of clear examples of nuclear-armed anocracies, the 1973 Arab-Israeli "Yom Kippur War" would appear to be the best choice. Tannenwald notes that Israel's refusal to sign the Non-Proliferation Treaty may reflect insensitivity (though not imperviousness) to the stigma associated with nuclear weapons, and that Israel's survival was arguably at stake in the 1973 war.[94]

HARD CASES IN A HARD FIELD

This article's purpose has been to make plain some of the ambiguities and contradictions within existing definitions and usage of LL/ML case studies. Researchers must be careful to account not only for the degree to which conditions countervail against theoretical expectations in a specific case, but their prior confidence in the theories being investigated for cases comparable to the one selected. Before using case studies to perform LL or ML tests researchers must devise theories with precise expectations and clear scope conditions. Rather than relying

[92] For more on the criteria for "decisive tests" within a Bayesian framework see Macartan Humphreys and Alan Jacobs, "Mixing Methods: A Bayesian Unification of Qualitative and Quantitative Approaches," paper presented at the American Political Science Association annual conference, Chicago, Illinois, September 2013; and the technical appendix in Andrew Bennett and Jeffrey Checkel, eds., *Process-Tracing: From Metaphor to Analytic Tool* (Cambridge: Cambridge University Press, forthcoming 2014).

[93] Research available at the time of Tannenwald's study provided strong evidence that mixed regimes are the most likely to take risky escalatory actions when they perceive themselves to be losing; Goemans, *War and Punishment*.

[94] Tannenwald, *The Nuclear Taboo*, 339, 379.

on ad hoc arguments, researchers should try to find reasonably established alternative explanations when specifying whether or not a case represents a difficult test. They must also take care that the selected case is not too easy or difficult for the theory of interest from the Bayesian perspective: can the case at hand significantly affect the researcher's confidence in a theory given the previous tests to which the theory has been subjected?

Even if researchers follow the preceding advice, they should remain cautious when applying the LL/ML labels. 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. As George and Bennett warn when invoking Bayesianism, "it is difficult to judge the probative value of a particular test relative to the weight of prior evidence behind an existing theory."[95] Aside from the complications that have already been covered, it is virtually impossible to know if the true explanation for a case's outcome is among the alternatives a researcher is considering.[96] If researchers are not mindful of the strong possibility that the actual explanation for the occurrence of a particular outcome is missing from the set of theories they are evaluating, they will tend to overestimate the degree to which theories have been validated by LL cases.

Such challenges are ubiquitous in political science regardless of a researcher's preferred methodology. Still, these issues loom large when researchers make bold claims, such as the ability to strongly support or undermine a theory's generalizability based on a single case. Fortunately, if the procedures laid out here can help researchers identify LL/ML cases, these tests are well suited for making the more cautious but still useful claim that, even if we cannot say that some theory is "true", it is better or worse as a general explanation for a political phenomenon

[95] George and Bennett, *Case Studies and Theory Development*, 120.
[96] Bennett, "Process Tracing," 713.

relative to alternative explanations. More hard thinking about hard and easy cases will help researchers in security studies to make such gains.
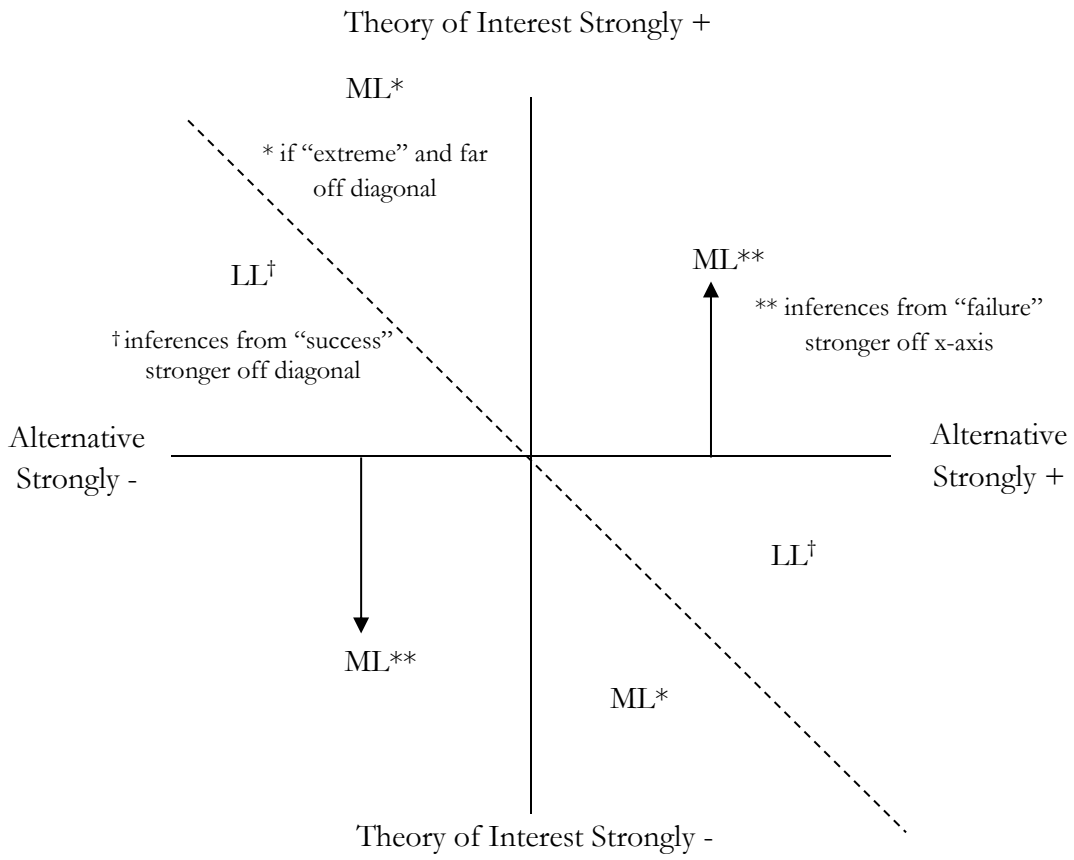
Figure 1: