Supplementary Materials for "Relationships among Rivals (RAR): A Framework for Analyzing Contending Hypotheses in Process-Tracing Research"

Appendix A: Insights from the Rival Space

While constructing a visual representation of the rival space is useful for any process-tracing approach, its utility is especially great in the explicit Bayesian context. This section illustrates a situation in which Bayes' rule can yield misleading results even when all estimated probabilities are correct *and* hypotheses satisfy mutual exclusivity. The problem begins when two hypotheses exhibit outcome mutual exclusivity, but not evidentiary mutual exclusivity (i.e. they make different predictions, but do not require different evidence). I demonstrate that if a piece of evidence is common under two competing hypotheses, yet rare overall, the standard incarnation of Bayes' theorem yields deceptively strong posteriors in favor of the wrong hypothesis. Modifying Bayes' rule with insights from the rival space can correct this problem.

To illustrate the problem, I draw on an epidemiological example precisely because they are often too straightforward to be effective illustrations of social-science research. Despite the accuracy of the probabilities and the likelihood that the patient suffers from only one disorder (i.e. mutual exclusivity of the outcomes), this example nonetheless exhibits the problem at hand, which should drive home the potential scale of the issue.

When Butterflies Deceive

A patient presents with a butterfly rash across the nose and cheeks. An eager resident doctor recalls that butterfly or *malar* rashes are a common in lupus; thus, she wants to test the probability of lupus (L) given that the patient presented with a malar rash (M), which is given by the following equation:

$$P(L|M) = \frac{P(L)P(M|L)}{P(L)P(M|L) + P(\neg L)P(M|\neg L)}. \tag{A.1}$$

$P(L)$ is our prior, defined here as the incidence of lupus in the general population:[1] 0.0047. $P(M|L)$ is the frequency with which patients afflicted with lupus present with

butterfly rashes: 0.55. $P(\neg L)$ is defined as $(1 - P(L))$, which represents the proportion of the population that does not have a lupus diagnosis. Finally, $P(M|\neg L)$ is a measure of how common malar rashes are overall: here, 0.033.[2] Thus, the probability the patient has lupus given a butterfly rash is,

$$P(L|M) = \frac{.0047(.55)}{.0047(.55) + .9953(.033)} = .073. \tag{A.2}$$

While the result is small in absolute terms, 0.073 is large relative to the prior, 0.0047. As such, observing a malar rash increases the doctor's confidence in a lupus diagnosis by a factor of 15.5. According to the Bayesian framework, a piece of evidence that is rare or unexpected overall should be given greater evidentiary weight—and indeed, both lupus and malar rashes are rare. This result might be interpreted as a smoking gun in favor of a lupus diagnosis based on the magnitude by which we updated.

However, when a piece of evidence is rare overall (i.e. contributes to a small denominator), researchers are at risk for too hastily eliminating alternative hypotheses on the basis of an artificially inflated posterior probability. Consider the case in which evidence K is rare overall (throughout the majority of Ω), yet common under both H_1 and H_2. If a researcher performed a Bayesian analysis on H_1 relative to all of Ω (i.e. the population at large), she would end up with a deceptively strong posterior.

Testing against the whole population can produce misleading results if two hypotheses rely on the same piece of evidence for their validation, yet that piece of evidence is rare outside of those hypotheses. Although malar rashes are rare *overall* (i.e. with an infinitesimal number of exceptions, malar rashes indicate either rosacea or lupus), they are common symptoms in both lupus and rosacea. As such, if we reduce the sample space to include just the hypotheses associated with butterfly rashes, we could then adjudicate among reasonable alternatives given the evidence. Since the likelihood of comorbidity between lupus and rosacea is approaching zero, these diagnoses are essentially mutually exclusive. Figure 1 illustrates this new rival space.

[Figure 1 about here.]

A Bayesian analysis on this new sample space reveals the sensitivity of Bayes' rule to properly defining relevant alternatives and understanding that a single piece of evidence might be observable under multiple alternatives (i.e. congruent). Lupus affects 1.5 million people per year and rosacea affects approximately 16 million.[3] Butterfly rashes manifest in 55% of lupus cases (825,000 people) and in 61.6% of rosacea cases (9,856,000 people). Thus, a total of 10,681,000 people are likely to exhibit this symptom. Testing the probability of lupus given a malar rash against the relevant sample space gives us,

$$P(L|M) = \frac{.0857(.55)}{.857(.55) + .9143(.616)} = .077. \tag{A.3}$$

The new posterior probability that the patient has lupus (.077) is *lower* than the prior (.0857) because this equation captures the relative frequency of diseases in which malar rashes are likely to occur. Thus, when a piece of evidence is expected under two competing hypotheses, but rare overall, Bayes' rule can produce misleading results in favor of any hypothesis tested its negation, which is the default input.

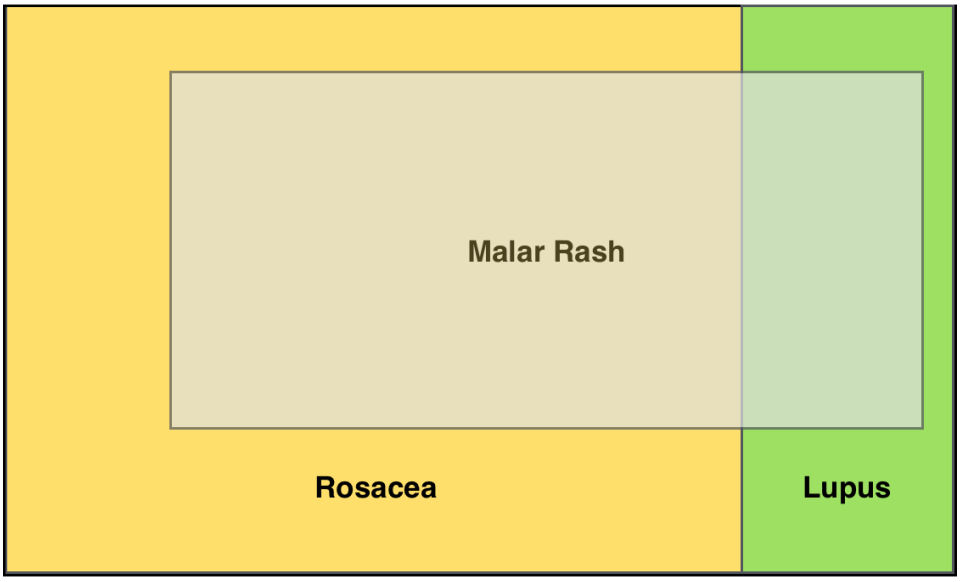

Figure 1: New Rival Space given Malar Rash

Appendix B: Using RAR to Broach the Nuclear Taboo

The second example I analyze brings the RAR framework to bear on Tannenwald's (2007) research on nuclear non-use. Tannenwald contends that in many cases, non-use is better explained by an emerging taboo against nuclear weapons than by rational deterrence theory. She convincingly establishes that deterrence alone cannot account for all instances of non-use, but she stops short of explaining the relationship between the two theories or indicating where one theory provides more analytic leverage than the other.

The RAR framework helps get traction on the ambiguities left in the wake of Tannenwald's study. This section establishes that the nuclear taboo and deterrence theory are not only congruent, but inclusive. I demonstrate the analytic purchase of this insight by integrating the taboo into a broader conception of deterrence theory and then using it to explain three additional anomalies in the historical record of the nuclear age that were heretofore outside the scope of her argument.

Main Hypothesis: Confirming the "Nuclear Taboo"

Tannenwald argues that the use of nuclear weapons in Japan prompted a global abhorrence to their destructive power. In the postwar period, this disgust coalesced into a set of norms she calls the "nuclear taboo." In this section, I present Tannenwald's hypotheses and propositions regarding the nuclear taboo and the evidence she offers to validate its role in U.S. decision-making.

H1: "[The taboo] is a necessary condition for the overall pattern of non-use" (2007, 54).

H1a Implication: No taboo was present when the U.S. used the bomb on Japan.

Since the taboo prevents use, Tannenwald first presents evidence that neither elites nor civilians viewed nuclear weapons as taboo during WWII. Transcripts from top U.S. officials suggest that no one at the time made a distinction between nuclear weapons and conventional bombs (2007, 79–80). On the civilian side, she cites a poll from *Fortune* magazine, which found that only "5 percent of the American public opposed use of the bomb under any circumstances," while "23 percent wanted to 'use many more of them before Japan had a chance to surrender'" (2007, 89).

H1b Implication: The taboo is present in instances of non-use.

Tannenwald considers numerous conflicts in which nuclear weapons were a strategic option (pre-Cold War, Korea, Vietnam, India-Pakistan crisis). In elite transcripts, she observes that the incidence of morally charged words like "revulsion," "repugnance," and "abhorrence" increased over time, thereby indicating a shift in the nuclear discourse from strategic concerns to moral concerns.

H2: Public pressure helped shape the nuclear taboo.

H2a Implication: public opinion shifted toward a stance of non-use.

To demonstrate shifting public opinion, Tannenwald contrasts the earlier Fortune poll with a 1954 Gallup poll, which reported that 76 percent of Americans did *not* agree that "we should go to war against Russia now while we still have the advantage in atomic and hydrogen weapons" (2007, 107-8). She also shows an increase in the number and membership of anti-nuclear civic associations both domestically and abroad by the late 1950s (158).

H2b Implication: Shifting public opinion had an effect on the elite decision-making.

In the wake of numerous grassroots organizations calling for bans on nuclear use and testing, Tannenwald provides evidence suggesting how this pressure manifested politically. Presidential candidates began incorporating test bans into their platforms (2007, 158), and the Soviet Union responded to the Third World non-aligned movement by calling for a test ban in 1955. Her evidence convincingly establishes that numerous campaigns "helped to stigmatize nuclear weapons" and delegitimize their use (2007, 162).

H3: If the taboo explanation is correct, the historical record should make explicit references to a norm or "taboo talk."[4]

Her evidence of "taboo talk" accumulates rapidly. In 1945 Secretary of State Dean Acheson wrote to Truman arguing that "the moral and political nature of our people is such that the use of the atomic bomb for an unwarned attack on another nation is not a practical possibility" (2007, 106).

Overall, Tannenwald provides compelling evidence that a nuclear taboo arose in the post-war years and played a crucial role in shaping elite decision-making. Yet, the strength of her overall argument could be enhanced with a more nuanced consideration of alternative hypotheses. Furthermore, while the following anomalies do not call into question the validity of her theory, readers interested in explaining the empirical record of the nuclear age more broadly may come away with three questions regarding the further applicability or limitations of the taboo: (1) How does one reconcile the taboo status of nuclear weapons with arms race and within-state proliferation? (2) How does one account for Western European states acquiring nuclear weapons *after* the taboo was clearly in place in those states? (3) If the taboo is as potent as Tannenwald suggests, how does one account for the intense fear of North Korea's nuclear arsenal and the same fear of Iran's potential acquisition?

Deterrence Theory: Establishing Congruence & Inclusion

The primary alternative explanation Tannenwald considers is deterrence theory. Deterrence, broadly defined, is "the act of dissuading an adversary from doing something it would otherwise want to do...through threats of either denial, or punishment, or a combination of these" (2007, 32). For deterrence to operate, the deterrers must satisfy three conditions: they must have "(1) credible capabilities, (2) a clearly communicated threat,

and (3) a credible willingness to carry out that threat" (2007, 31). Its popularity, parsimony, and accurate predictions notwithstanding, Tannenwald argues that deterrence theory's *explanatory* power is limited. She cites four major empirical anomalies that call into question the scope of circumstances for which deterrence theory can account, and from there, she argues for a shift in attention to the nuclear taboo.[5]

Based on the empirical anomalies and her evidence in favor of normative influence on nuclear decision-making, Tannenwald concludes that "there are a significant number of cases of non-use in which deterrence simply did not operate" (2007, 33) and "important causal factors lie outside deterrence theory" (2007, 36). These claims entrust Tannenwald with a twofold responsibility: first, she must show that the other factors matter, which she does; second, she must show that these factors indeed lie outside deterrence theory, which she does not. Furthermore, her analysis admits of a disparity between the theory she originally introduces and the version of deterrence theory she actually tests. Though Tannenwald defines deterrence broadly, she notes that many scholarly examinations of non-use stem from the neorealist tradition, which subscribes to a narrower conception known as rational deterrence theory (hereafter, RDT) (2007, 31). RDT includes numerous simplifying assumptions about what drives decision-making: utility maximization, materialist cost-benefit logic, and strong rationalism. Her empirical tests focus only on evaluating this version of deterrence, yet her conclusions about its inadequacy are directed at the deterrence explanation as a whole. She tests RDT with the following analysis, which implies evidentiary mutual exclusivity between deterrence and the taboo.

H1: Under a "purely materialist explanation" decision making would reflect cost-benefit thinking (2007, 51).

> She argues that even within the Eisenhower administration, which considered deploying nuclear weapons on multiple occasions, she is left empty-handed. Among long lists of potential detractors, she observes, "notably absent...is any clear statement that use of atomic weapons would risk provoking global war with the Soviet Union, a deterrence concern" (2007, 145). Tannenwald, however, interprets an absence of corroborating evidence as evidence of absence.

H2: "Decision-making would not reflect *any* taboo factors" (2007, 51).

> As the previous section depicts, Tannenwald finds extensive evidence that normative factors played a vital role in leaders' calculus. She thus interprets normative discourse as undermining deterrence theory.

The RAR framework could have enhanced the rigor of her analysis on two fronts. First, she dismisses deterrence on the basis of a assumption in one version of deterrence theory—materialist considerations did not drive non-use. Although RDT is built on the assumption that deterrence operates only through materialist channels, and her evidence of taboo talk suggests something other than material concerns factored into non-use, Tannenwald overextends the implication of this false assumption by relegating

deterrence theory to the sidelines rather than searching for opportunities to modify the theory. This error stems from conflating one channel by which deterrence works (materialist concerns) with a deterrence explanation more broadly, which could operate along numerous dimensions.

Second, dismissing the deterrence explanation was made easier by the absence of a crisp framework for modeling the relationship between deterrence and the nuclear taboo. Indeed, Tannenwald's discussions of how the two relate vacillate between implying mutual exclusivity at some points,[6] to implying congruence[7] at others. This ambiguity both undermines her analysis (since her tests imply mutual exclusivity) and leads her to overlook a major contribution her theory could have made to the deterrence framework. Here, I examine the relationship between the two theories and show that the taboo might be one of many possible dimensions along which deterrence operates, thereby rendering the theories not just congruent, but inclusive.

Establishing Congruence and Inclusiveness

Q1: Do deterrence theory and the nuclear taboo generate divergent predictions?

> No. Neither theory would expect non-use in WWII because (1) the U.S. had a nuclear monopoly, and (2) the taboo was virtually nonexistent. Neither would predict use during the nuclear age. The only ambiguous point is during the post-war U.S. nuclear monopoly. However, non-use is puzzling under *both* theories here because the norm was not yet entrenched, but the USSR had not yet acquired nuclear capabilities to act as a deterrent. Alternately, if we consider the theories in *conjunction*, non-use could be explained by a combination of a nascent taboo plus the USSR's conventional military superiority.[8] Thus, outcome mutual exclusivity is especially unlikely.

Q2: Does the evidence in favor of one rule out the other?

> No. Evidence in favor of RDT would indicate material concerns; evidence in favor of the taboo would indicate normative concerns. Nothing, however, prevents material and normative concerns from being in play simultaneously.

Q3: Are the two theories candidates for inclusiveness?

> Yes. After demonstrating that the materialist assumption of RDT does not hold, the taboo theory can be recast as a novel dimension of deterrence theory that fills in this gap. Deterrence requires behavioral modification in the face of a credible threat. The nuclear taboo represents a novel *mechanism* or motivator for deterrence (in addition to material concerns), and the incensed public represents a novel *channel* through which deterrence can operate (since the U.S. public can levy a credible threat of ousting leaders). As such, a key requirement for public normative concerns to operate as a mechanism of deterrence is for the public to have some power over leaders, thereby suggesting that we should only observe this dimension of deterrence in democratic states.

The RAR framework reveals that the taboo is a good candidate for inclusiveness under deterrence theory. This insight has three positive implications for both Tannenwald's research and our broader understanding of nuclear non-use. First, integration makes a major contribution to deterrence theory. Recasting the taboo as an extension to deterrence adds nuance to the theory by revealing a variety of channels through which deterrence can operate. More specifically, this expansion reveals that the potential target in a conflict and the source of the deterrent threat need not be the same actor and the nature of the deterrent threat need not be of the same type.[9]

Following from the first, the second benefit of integration is a more comprehensive framework for studying deterrence. To explain deterrence in a given case, researchers would first need to identify the possible channels and sources of deterrence. They could then examine the historical record to test which channel (whether material, normative, or otherwise) was most salient and which source(s) (deterrence arising from the target, peers, or the public) leaders felt the deterrent threat from.

The third benefit of an inclusive framework is that it resolves the anomalies left in the wake of Tannenwald's study. Both the nuclear arms race and proliferation can be explained by appealing to a *different dimension* of deterrence: perhaps material and safety concerns outweighed any taboo of *having* nuclear weapons, just not the taboo of *using* nuclear weapons. Additionally, an integrated theory explains why the taboo is expected to be more salient in some countries than others: namely, why (1) "U.S. political leaders consistently rejected [preemptive nuclear strikes]" on the basis of the taboo, "despite the fact that they believed no moral considerations would limit the Soviets from launching an aggressive war" (2007, 106); and (2) the U.S. and many other states express grave concern over North Korea's arsenal and Iran's potential acquisition. If deterrence is only expected to operate when the deterring party can make a credible threat, we should expect the public's normative concerns to manifest as deterrent threats only in democratic states, in which leaders are accountable. Thus, *some* deterrence likely operated in the Soviet Union, but perhaps it was only on a material dimension. States that lack institutions for public onion to sway elite decisions are essentially missing this large force pushing for non-use.

References

Tannenwald, Nina. 2007. *The Nuclear Taboo: The United States and the Non-Use of Nuclear Weapons Since 1945*. New York: Cambridge University Press.