Multi-Method Research: The Case for Formal Theory

Hein Goemans
University of Rochester

William Spaniel*
University of Rochester

September 27, 2013

Abstract

We argue that formal theory and historical case studies, in particular those that use process-tracing, are extremely well suited companions in multi-method research. To bolster future research employing both case studies and formal theory, we suggest some best practices as well as some (common) pitfalls to avoid.

Keywords: *Formal theory, process tracing, case studies, multi-methods.*

*We thank Brenton Kenkel and the participants at the History, Method, and the Future of Security Studies Miller Center of Public Affairs conference at the University of Virginia, September 26-27, 2013 for their comments.

Scholars in International Relations in general and in Security Studies in particular use a variety of methods to communicate their ideas and persuade the reader of the value of those ideas. It seems that in the late nineties, a critical number of scholars concluded that the best way to communicate and persuade was by a combination of several methods. Many now consider such multi-method research as a standard to be emulated. Such multi-method scholarship typically employs one method for theory (construction) and another for empirical evaluation. Rationalist theories, our main focus here, are then most often paired with statistical analyses, occasionally with historical case studies (Slantchev 2002; Goemans and Fey, 2011; Debs and Monteiro, 2014), and recently also by some experiments (Walter and Tingley, 2012; Dolan, 2013).[1]

For at least a decade thus, young scholars aim to first build formal models of strategic interaction and then statistically evaluate (some) equilibrium predictions, which turns out to be less than straightforward. Signorino (1999) conclusively demonstrated the perils of estimating models of strategic interaction without statistically correcting for the *non*-independence of the observations. If the model describes a data-generating process, the empirical model must be able to retrieve it. Because a proper statistical specification of the model is so difficult, researchers sometimes rely on verbal reasoning - often arguing for un-modeled selection effects - admitting in effect that their model only captures part of the data-generating process (Bueno de Mesquita 1999; but see Fearon (2002) who deduces a selection-effect to explain some puzzles raised by Huth and Russett (1990) and Lebow and Stein (1990)). They then present a series of statistical models which purport to measure the moving parts of the formal model against the observed data.

But as has been acknowledged for quite some time, the strategic statistical models pro-

[1] In an interesting new paper, Dolan (2013) combines a formal model with insights from psychology and uses an experimental design to evaluate his theory.

posed to deal with these problems suffer from some flaws of their own, principally in that they assume the strategic model is the true and only data generating process. While many seem to think that "people who use math" are natural scholarly partners with"other people who use math" it is clear by now that scholars with a quantitative empirical approach face a pretty difficult task contributing to scholars who are primarily concerned with building formal theory.[2] Overcoming these difficulties has been a major focus of sustained interest in the last decade or more for scholars in quantitative research (most notably the hot area of 'causal inference').[3]

While there has been some progress in integrating statistics with formal models, little attention has been paid to the integration of formal models and case studies. This is odd, we would argue, because targeted historical research in case studies and process tracing is well suited, and arguably better suited, to evaluate the empirical usefulness of formal models. We would be amiss, however, if we did not begin by pointing out that a reading of history is often the basis for formal research. The reader is struck by a note, an anecdote, or some pattern depicted by a skillful historian and starts to wonder how the event or chain of events came about. The fundamental puzzle, thus, came from a case study. To use that same case study as evidence or illustration of the model is of course disingenuous - it is effectively selecting on the dependent variable – but publishing norms in our field seem to frown upon papers that begin with a historical puzzle and end with a formal exploration of that puzzle.[4] Note that it is often the very richness of the historical case study which makes the puzzle a good candidate for formal exploration. In turn, we argue, it is the very richness of the historical record which often makes historical case studies the best tool to evaluate the plausibility of a formal model. The simple point is that only process-tracing can really

[2]Signorino and Kenkel (2013), however, bring a new suite of tools to facilitate such cooperation.

[3]Imai et al., "Unpacking the Black Box of Causality," in the *APSR*.

[4]We would argue that such papers should be allowed to flourish and 'inject' new theory into the discipline so that other scholars can manipulate it, criticize it, and, yes, test it as they see fit. That does constitute a scientifically legitimate project.

examine whether decision-makers were thinking strategically. We can often overhear, or trace a decision maker considering her opponents calculus and find evidence that a decision maker chose one course of action over another because the other course of action was "of the equilibrium path." Moreover, we would argue, historical case studies can make it possible to match the explanatory theoretical variables to the real world very precisely, often much more accurately than with a large data set with many cases. This is extremely attractive from a formal perspective as formal models present a very precise and specific set of variables, (values of) parameters and their relationship. Careful historical research can spend the time to delve into cases and offers a much more fine-grained, and therefore potentially much closer, empirical match with the theoretical concepts and parameters of interest in the model.

A common complaint of formal models is that the assumptions that drive the model often don't hold "universally." There are two important things to keep in mind here. First, "assumptions" should be more fruitfully thought of as "boundaries conditions" that define the scope of cases two which the argument applies. Second, and as a result, if the scope is quite narrow and the number of applicable cases quite small but very important, statistical inference may be very weak tool, and case studies that trace the model may give us a huge gain in understanding. (Think the Big Bang, perhaps?)

Another powerful advantage of complementing a formal model with historical case studies follows from the potential of a close match between theoretical and empirical parameters: a good writer can then let the empirical case study re-tell the theoretical story. This is a highly effective rhetorical strategy, and indeed the main case study strategy we propose. It means scholars must take the formal model, historical research and the connection very serious. These advantages and potential complementarities of case studies seem oddly overlooked by many formal modelers. To change this attitude, we suggest some avenues to fruitfully explore formal models in historical case studies. To build a case for combining formal models with case studies, we first present a simple model. Next we highlight some common mistakes

made by formal modelers who attempt process tracing.

1 A Simple Model to Make the Case

Every game specifies the actors, the actions that the actors consider, the actors' beliefs and their pay-offs. The analyst chooses some fixed elements – typically the number of actors, for example – and chooses to *parameterize* the payoff functions depending on the goal of the model.[5] To properly trace a formal model, we must trace all of these. This is necessary not just to gain maximum leverage, it is required to evaluate the *causal mechanism* proposed by the model. A causal mechanism is not the equilibrium outcome for a *particular* parameter space that interests the researcher. Rather, the mechanism is the *entire* equilibrium. This means that a causal mechanism actually describes a logical progression from variation in the model's parameters to variation in outcomes. The case study researcher thus also needs to be able to describe what the outcome *would have been* in a world with slightly different parameters, and – getting ahead of ourselves – even including sometimes different "off-the-equilibrium-path" beliefs.

To illustrate some basic strategies and pitfalls of a case study trace of a formal model, we require ... a model. We choose the classic preventive war model (Fearon 1995) for its simplicity and familiarity. Two states, R(iseland) and D(eclineland) bargain over a pie sized 1 in the shadow of shifting power and war. The interaction occurs over two periods. In the first period, D offers $x \in [0, 1]$ or prevents. If D makes the offer, R accepts or rejects. If R accepts, the interaction advances to the second period. D then offers $y \in [0, 1]$, which R accepts or rejects. This is a game of complete information, where each actor knows what the other knows.[6]

[5] As noted above, we can think of the fixed elements of the model as its boundary conditions.

[6] Once we consider models of incomplete information with perfect Bayesian equilibria, it becomes significantly more difficult to process trace how beliefs change, although we might still be able to established that indeed they did.

Payoffs are as follows. States share a common discount factor $\delta \in (0,1)$. In other words, states each value the future compared to the present the same; they are identically myopic or patient. If peace prevails throughout, R receives $(1-\delta)x + \delta y$ (a combination of the discounted values of the first and second offers) and D receives $(1-\delta)(1-x) + \delta(1-y)$. If D prevents or R rejects in the first period, R earns $p - c_R$ and D receives $1 - p - c_D$, where $p \in [0,1)$ is the probability of victory and $c_i > 0$ are each state's costs of war. Finally, if R accepts in the first period and rejects in the second period, R receives $(1-\delta)x + \delta(p' - c_R)$ and D receives $(1-\delta)(1-x) + \delta(1-p'-c_D)$, where $p' \in (p,1]$ is the probability of victory after the shift, in the second period.

Proposition 1. *The preventative war game has a unique subgame perfect equilibrium. War occurs on the equilibrium path if and only if* $p' > p + \frac{(1-\delta)p + c_D + \delta c_R}{\delta}$.

The proof is trivial but the intuition is as follows. In the second period, D offers $y = p' - c_R$ and R accepts. Thus, at most, D can receive $(1-\delta) + \delta(1-p'+c_R)$ if it does not prevent. This is less than the $1 - p - c_D$ D receives if it prevents if $p' < \frac{p + c_D + \delta c_R}{\delta}$, which is the cut point from the proposition.

Figure 1 presents in graphic form the *cutpoints* for the combination of values of two of the main parameters of the model, holding p, c_R and δ constant. Note that the precise location of the cutpoint line, and therefore the 'peace area' and 'preventive war area' changes depending on the value of those parameters.

2 Tracing the Model

Parameters

Empirically tracing a model is not without challenges. It is easy to select on the dependent variable. In an attempt to illustrate preventive war, a researcher might find an instance

in which Riseland was growing more powerful and Declineland declared war. Despite the similarities, this is *not* reflective of the model. Rather, it reflects the model's outcome–which may have been the result of the causal mechanism of the model or may have been the result of a different mechanism altogether. Focusing on the outcome ignores the process, which defeats the purpose of modeling.

In contrast, proper process trading requires a through integration with the model. Although researchers pay great costs in terms of time and page length to develop models, case studies all too often immediately abandon the model's inputs. In contrast, we argue that any effective operationalization of a model must pay heed to its parameters. Who are the actors? What are their actions? Their beliefs? The parameters that go into their payoffs? These are a model's independent variables. Without an exploration of them, we cannot discern which outcome the equilibrium will produce.

Relating this to the preventive war model, the first three parameters are *fixed* and relatively straightforward to operationalize by design. The actors are states; we can trace their behavior by focusing on the decision-making apparatus of the state as for example done by cabinets, the president, or the autocrat. Since we have assumed complete information, we must pick cases where both sides know of an impending shift in the probability of victory; if the declining state learns of the power shift only after it fought a war for other reasons, the mechanism was not in play. Such a shift in the probability of victory could be and has historically been driven by a variety of factors, hence we could focus on Germany and Russia in the beginning of the twentieth century, or we could look at the US and the Soviet Union in the late 80s and early 90s or to Iraq in the early 2000s (Spaniel 2013; Debs and Monteiro 2014).[7] We also have assumed states have a very restricted set of possible actions: make an offer, launch (preventive) war, accept or reject the offer. A declining state that could

[7]Note that if we pursued another model, such as Debs and Monteiro (2014) that argues for the role of imperfect information, we must deal with an additional layer of historical complexity.

"easily" or "cheaply" get support from a potential ally would therefore be a relatively poor historical match to trace the model.

Most of the empirical action, unremarkably, lies in the *parameterization* of the payoffs. From the model, states have probabilities of victory, costs of fighting, and valuations of the future.

The central advice is for researchers to find ***cases where the historical evidence allows one to evaluate counterfactuals***. That requires work.

Our model includes one parameter that would seem very difficult to historically trace: the states' discount factor, which is often deemed unobservable.[8] Note, however, that we assume the states have the same discount factor. A careful historical trace would thus have to compare and evaluate whether the two contending states could plausibly be argued to have significantly different discount factors. It might be possible to see consistently more myopic policies by one state, or it might be the case that the main decision maker is old and terminally ill – much like Walt in *Breaking Bad* – and therefore cares much more about the near future than the longer term, which he does not expect to experience. If historical research suggests that states indeed have significantly different discount factors, then the case is not well suited to trace the model. Note, moreover, that such preliminary research on states' attitudes to the future is also necessary to provide some historical basis for the judgment whether the common discount factor was high, low or somewhere in the middle.

Some parameters are easier to evaluate than others. Costs of war, likelihood of victory, cost of economic sanctions, and gains from trade can all be quantified. Unobservable parameters, such as patience (a discount factor) and uncertainty (prior beliefs) are not as easy to analyze.

[8]The discount factor is a parameter in the model, but the fact that it is the same for both players is *fixed*.

Avoiding Formal Fallacies: Interpreting Cutpoints

Cutpoints are not easy to interpret. Most models produce complex interaction terms as cutpoints for equilibrium outcomes. Case study analysis must accordingly interpret the *entire* cutpoint. Let us put forth an often overlooked, major advantage of a case study trace of formal models. Cutpoints make *deterministic* statements, whereas quantitatively we typically engage in *probabilistic* analysis. Quantitative analysts then rely on a sleight of hand here of the following kind: adding small amounts of noise or measurement error turns the hard cutpoint would turn into a soft probabilistic threshold. Case studies, in contrast, can be finer-grained

Broadly, there are two ways to do this. The first is to organize a cutpoint in a manner that has a natural informal interpretation. Although the proposition above specifies the cutpoint for preventive war as $p' > \frac{p+c_D+\delta c_R}{\delta}$, recall that it is derived from $1-p-c_D > 1-\delta+\delta(1-p'+c_R)$. The left side of the inequality represents D's payoff for preventive war in the first period. The right side of the inequality represents Declineland's payoff for taking everything in the first period plus the remainder of the good after accommodating Riseland in the second period. Such a natural interpretation allows a researcher to look into Declineland's history and the diplomatic archives to check whether policymakers at the time believed that the future concessions it would have to make justified the costs of preventive war.

Unfortunately, some cutpoints lack an easy informal interpretation. For these, researchers must evaluate individual parameters. However, the researcher must not forget that the individual parameters are meaningless without the context of the cutpoint as a whole. For example, if preventive war occurs when $p' > \frac{p+c_D+\delta c_R}{\delta}$, a researcher might attempt to show the rationality of preventive war by analyzing Declineland's cost for fighting. Perhaps he finds that advances in military technology and the burden of keeping the peace meant that c_D was very close to 0. He may then conclude that the causal mechanism explains Declineland's

preventive action.

However, given that information alone, there is insufficient evidence to conclude whether the causal mechanism plausibly explains the observed phenomenon. Note that even if $c_D = 0$, making that substitution into the cutpoint yields $p' > \frac{p+\delta c_R}{\delta}$. Consequently, the causal mechanism dictates that preventive war will not occur if c_R is sufficiently large. The author's mistake was forgetting that c_D works in conjunction with other parameters for that cutpoint.

Although the safe path is to always analyze *all* parameters within a cutpoint, sometimes this is unnecessary. Recall that p', at most, equals 1. As such, if the right side of the inequality is greater than 1, the causal mechanism says that peace prevails. Because all values on the right side of the inequality are positive, a single large value from one parameter (like $c_D > 1$) can carry the whole term over that threshold.[9] So while low values of c_D cannot prove that the causal mechanism correctly predicts war, high values of c_D can prove that the causal mechanism correctly predicts peace.

Counterfactual Analysis

A causal mechanism is not an outcome; rather, the mechanism is the *entire* equilibrium. It is therefore insufficient to look only at the realized parameters and outcomes to validate the model's prediction. We must instead think seriously about the counterfactual in terms of slightly different realities and off-the-equilibrium-path behaviors.

To begin, consider Figure 1, which takes the extent of the power shift and Declineland's cost of war as inputs and produces an equilibrium prediction. When costs are low and the shift is great, as in the top triangle, war results. Otherwise, peace prevails. Parameterizing the independent variables may lead a researcher to conclude that the observed case falls in that triangle, as diagramed. But because the mechanism is the entire equilibrium, peace would need to be the result in slightly different counterfactual worlds in which the power

[9] This is because $\delta \in (0, 1)$, meaning that any value for δ increases the value of the numerator.

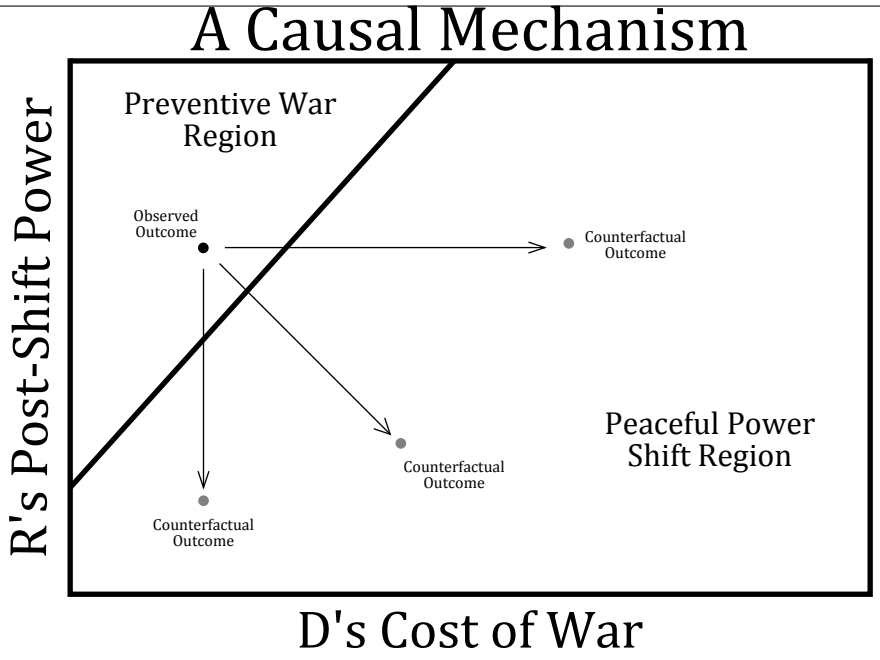

Figure 1: Equilibrium outcomes for a given value of p' and c_D. Remaining parameters are set to $p = .1$, $c_R = .09$, and $\delta = .9$. War only occurs when the power shift is great and D's cost of war is low.

shifts are smaller and the costs of war are greater.[10]

This, of course, means the author must use counterfactual reasoning to analyze the application of the causal mechanism. There are three ways to accomplish this. First, the researcher can look for an exogenous shock that altered the relevant parameters at hand. Perhaps previously Declineland worried that military attrition in a war with Riseland would be high and Thirdland would extract concessions from Declineland should it attempt a preventive strike. This functionally inflates the value of c_D. But suppose that Thirdland recently had an economic collapse and no longer poses a military threat to the international system. Then c_D shifts to a lower value. If Declineland did not prevent previously but did initiate a conflict after the fall of Thirdland, the situation matches the causal mechanism.

[10]For a related argument, see Bruce Bueno de Mesquita, "Counterfactuals and International Affairs: Some Insights from Game Theory," in Philip E. Tetlock & Aaron Belkin, *Counterfactual thought experiments in world politics; logical, methodological and psychological perspectives*, Princeton, NJ: Princeton University Press, 1996.

This is the most desirable counterfactual, as it relies the least on the researcher's ability to make historical inferences.

Second, the comparative method has long suggested using *two* cases to serve this purpose. The prescription here is to find two cases that share all relevant independent variables of interest except one and then show that the dissimilar independent variable must account for the variation in dependent variables. For example, Declineland begins a preventive war against Riseland's nuclear weapons program but not against Assentland's tank program; all parameters between Riseland and Assentland are the same, except Riseland's p' is greater due to the greater coercive strength of nuclear weapons. Here, each situation is the counterfactual for the other.

Third, the researcher can construct a counterfactual using a plausible alternative historical narrative. For example, imagine Declineland initiated against Riseland. Then for the causal mechanism to hold, it must be true that Declineland would not have fought in a counterfactual world in which Riseland owned nuclear weapons and could have delivered a countervalue strike such that c_D exceeded 1. This is the least desirable of the three options, as it requires significantly more work than the other two options. It is also less easily replicated as it requires the author to make informed decisions about how actors would behave in the constructed counterfactual world. However, sometimes case studies do not provide any better alternatives.[11]

Similar logic applies to off-the-equilibrium-path actions.[12] Consider a situation in which Riseland is slowly growing more powerful and Declineland declares war. Due to the known

[11] If, for example, one argues that regime type is crucially important and that Mixed regimes behave differently than Democracies, it might be possible to trace what the preferences of the Mixed Regime would have been if it counter-factually had been a Democracy. Goemans (2000) found a nice example during WW I where the German Parliament adopted a resolution which fundamentally differed from the policy of the leaders of the Mixed regime and allowed him to plausibly infer what a Democratic Germany's policy choice (offer) would have been.

[12] See Barry R. Weingast, "Off-the-path; A game theoretic approach to counterfactuals and its implications for political and historical analyses," in Tetlock & Belkin.

power shift, it may be tempting to label this as a preventive war. However, the causal mechanism explicitly states that Declineland will reach a disadvantageous peace in the second period. If the states could contract a distribution more favorable to Declineland (that is, if the states could resolve the commitment problem), then war would not happen. The researcher must consequently investigate what Declineland's leaders believed would be the outcome of a post power shift world and confirm that the off-the-path expectation matches the empirical belief.

Similarly, it must be the case that Declineland would not have declared war later on. Otherwise, we would be explaining war today by pointing to war tomorrow. This is problematic because other causal mechanisms–perhaps information problems or preemptive war–actually explain the observed outcome, leading the case study to reach a false conclusion. Consequently, case studies with outcomes that fail to reach the end of the interaction require counterfactual analysis to ensure that actors would behave in the theorized manner to in turn rationalize observed outcomes.

3 Conclusion

Notes: the better the empirical fit of concepts and parameters of the model and the historical process tracing, the more confidence that you are actually capturing the mechanism historically. Analytical narratives. If you don't clearly operationalize your independent variables, you are always opening yourself up to selecting on the dependent variable