H-Diplo | Robert Jervis International Security Studies Forum

Roundtable 14-16

Teaching Formal Theory in International Relations

31 March 2023 |PDF: https://issforum.org/to/jrt14-16 | Website: rjissf.org

Editor: Diane Labrosse | Commissioning Editor: Andrew Szarejko | Production Editor: Christopher Ball

Contents

Introduction by Michael A. Allen, Boise State University

> "Is there any way you will trust me?" Inigo Montoya asks a masked man below him who is free-climbing the rock face of a mountain. Someone had hired Inigo to kill the Man in Black. He is eagerly waiting to fight him in a duel, but the Man in Black believes that if he takes the rope that Inigo has offered him, Inigo will take advantage of the situation, cut the rope, and kill him.
>
> "Nothing comes to mind," the Man in Black quickly responds as he strains to hold his position.
>
> Inigo offers, "I swear on the soul of my father, Domingo Montoya, you will reach the top alive."
>
> Satisfied that Inigo has found an option that credibly commits him to the deal, the Man in Black finalizes the bargaining with "throw me the rope." Inigo upholds his end of the deal, and the two, after a rest, proceed to their duel.[1]

Like the four contributors in this volume, I weave formal modeling throughout my classes. Whether teaching fundamental international relations, civil war and terrorism, or graduate courses on conflict, formal models can often give us a common vocabulary to understand assumptions, frameworks, and approaches to international affairs. Even in my 100-level course on individual leadership for first-year students, I let my students know that we need to understand how people make decisions to lead effectively.

The duel between the Man in Black and Inigo sets up a good conversation with my students about establishing when communication does or does not matter and modeling such moves in a formal model. The scene continues beyond the above interaction to set up two more immediate puzzles. First, both fighters begin the duel with their non-dominant hand unbeknownst to their opponent. Why would a sword fighter fight with a non-dominant hand against an opponent? Is there an advantage to fighting with less than your full capacity? Second, at the end of the duel, Inigo loses and prepares to be killed by the Man in Black, but the Man in Black decides to knock him out instead. What considerations might someone face in deciding to spare or kill an opponent? As the contribution from Xiaoli Guo suggests, I use clips from movies like *Golden Balls*, *The Dark Knight*, *Footloose*, and other cultural moments to create a common ground for my students to engage, debate what the proper decision is, and construct models that contain the essential elements of the clips we watch. After doing so, we analyze the conclusions that the models lead us to make.

An essential facet of formal models is that we are explicit about our assumptions. Every model or theoretical argument has assumptions. In the non-formal world, those assumptions can be hidden or transparent. Formal modeling leaves little wiggle room as you assign variables, values, actions, preferences, and strategies to a scenario. This process does a few wonderful things that can enhance a discussion in class.

First, being explicit has the benefit of having a student (or an instructor) think through their argument and make it available for others to digest. Hein Goemans' article discusses the virtues of teaching students how to organize their arguments and providing them with knowledge that can ripple through their careers and enhance their understanding to avoid the traps of common misunderstandings. From a discipline-building perspective, Bear Braumoeller notes that the scarcity of theorization in international relations and formal modeling allows us to recover what we have lost from other disciplines. Second, we can create a fictional world where those assumptions are true. We pretend "as if" the assumptions of a model are valid and allow

[1] Rob Reiner, *The Princess Bride.* Act III Communications, 1987

the model to help us find what the world looks like when the math finalizes. While the models are a way to simulate and understand reality, they also take something from a thinker's mind and build a world from those thoughts that we can test out.

Third, and one of my favorite ideas, is that a formal model, once articulated, allows us to assess the counterfactual. Much like science fiction allows us to ask "what if" and examine the conclusions a writer or director has made, formal modeling allows us to access the logic of a "what if" question in a method that is translatable to other students, teachers, and scholars.[2] Scott Wolford's contribution discusses how he uses formal modeling to teach World War I in "real-time." Each model and decision presents its what-if as students weigh decisions made during World War I and consider the possibility of different choices.

The writers in this forum have formally modeled key processes in international relations by asking questions like how great powers shape the system to their constituents' demands and how that produces conflict with other great powers.[3] How does the rewarding of risky behavior by selectorates affect leaders' willingness to engage in war?[4] How does learning from other states affect their performance during the COVID-19 pandemic?[5] If new leaders must prove their resolve internationally and rivals want to test their resolve, how does that affect war propensity?[6]

Each of these questions posed by the contributors in their research offers a model of our world. However, the models also allow us to ask counterfactual questions: What if a state's ability to influence the international system does not match its relative capabilities? What if selectorates do not care about foreign policy? What if states did not model the behavior of other states during the COVID-19 pandemic? How would conflict bargaining between states change if a leader's willingness to go to war over an issue was public information? Better than any science fiction work, formal models allow us (perhaps more importantly, our students) to envision numerous alternative scenarios (and whole universes!) just by altering an assumption or changing a variable.

While the power of formal modeling can be a good sell to students, math can be daunting. While teaching methods, I often espouse the belief that math, just like writing, takes practice.[7] Even people who are exceptionally gifted at understanding math will arrive at points where the math takes work. Such a barrier, for all skill levels, poses the question of whether we work and remove that barrier or whether we decide that this is a barrier we will not cross. One way to make hard work of math more palatable is by weaving in topics they care about, everyday situations they may encounter, or popular culture they are consuming. While my dated example of finding a credible commitment in the *Princess Bride* heads this article, popular culture will continue to provide game theoretic examples in film, shows, reality television, music, and other venues. By engaging in contemporary mass culture, the way we discuss formal models can create an intellectual symbiosis where understanding an interest can help us understand a formal model and vice versa.

---

[2] Michael A. Allen and Justin S. Vaughn, "Introduction: Science Fiction as a Tool of Political Science," in Allen and Justin S. Vaughn, *Poli Sci Fi: An Introduction to Political Science through Science Fiction* (Abingdon, UK: Routledge, 2016): 1-8. Routledge.

[3] Bear Braumoeller, "Systemic Politics and the Origins of Great Power Rivalry," *American Political Science Review* 102:1 (2008): 77-93.

[4] Hein E. Goemans and Mark Fey, "Risky but Rational: War as an Institutionally Induced Gamble," *The Journal of Politics* 71:1 (2009): 35-54.

[5] Xiaoli Guo and Dmitry Ryvkin, "When is Intergroup Herding Beneficial?," *Mathematical Social Sciences* 120 (2022): 66-77.

[6] Scott Wolford, "The Turnover Trap: New Leaders, Reputation, and International Conflict," *American Journal of Political Science* 51:4 (2007): 772-788.

[7] Hermundur Sigmundsson, R. C. J. Polman, and H. Lorås, "Exploring Individual Differences in Children's Mathematical Skills: A Correlational and Dimensional Approach," *Psychological Reports* 113:1 (2013): 23-30.

This volume offers four complamentary perspectives on bringing formal models to the classroom.

Bear Braumoeller reflects upon the aversion to teaching formal models in the social sciences. While international relations and foreign policy analysis are data-rich, the theorization that drives the empirical testing of our arguments is lacking. Echoing Georg Box's fundamentally important argument that "all models are wrong, but some are useful,"[8] Braumoeller conveys the need for our students, who will be future scholars, to have a way to convey their theories. After all, all of the advanced statistical models we use and how we construct our data imply theoretical arguments. However, when those assumptions and arguments are left unstated at the theoretical stage, we no longer have a connection between what we are testing and our understanding of how the world works. Unlike earlier generations of scholars, we have more tools to build our theoretical models and fewer excuses not to use them.

Hein Goemans thinks critically about the organization and structure of our knowledge. Students, despite decades of education and vast amounts of knowledge, often lack the tools to organize their knowledge and translate those knowledge structures to others. Goemans breaks down the components of formal models to his students and relates how those building blocks organize what we value in explaining social phenomena. Once we have those tools, we can better understand even fundamental ideas, like the Prisoner's Dilemma, why many sources get it wrong, and how we can elaborate on our foundational model to build additional insights.

Xiaoli Guo connects the game theory we use in international relations to our daily decisions. Building from scenarios that they are familiar with or that an instructor can easily show in the classroom, students can begin to understand concepts of rationality and more complex decisions such as alliance behavior or arms races. Tweaking the games can offer additional insights, new perspectives from different kinds of theories, or reveal altruism or other factors where we would not expect them.

Scott Wolford's essay discusses how he sought to bridge the lessons from international relations scholarship with understanding international relations. This quest led to a revolutionary rework of two classes into one. His contribution takes us through the course "World War I in Real Time" as students encounter puzzles as they locate themselves in the First World War, build models to identify the core issues, and then solve those problems. While there existed barriers to such a class, namely in math and reviewers, the payoff to students and their comprehension proved quite large. Thankfully, Wolford has produced a textbook for others wishing to teach a similar course.[9]

The four contributions to this roundtable offer us practical, pedagogical, and disciplinary insights about formal modeling in the classroom. From how to pair it effectively with content to the long-term benefits to a student's and scholar's career, the authors in this roundtable make a case for a more comprehensive implementation of one of our discipline's core tools.

Participants:

Michael A. Allen is a Professor of Political Science in the School of Public Service at Boise State University. He recently published *Beyond the Wire: US Military Deployments and Host Country Public Opinion* with Oxford University Press (October 2022). His work focuses on conflict, asymmetry, and foreign policy with a

[8] George E.P. Box, "Robustness in the Strategy of Scientific Model Building," in *Robustness in Statistics* (Cambridge, MA: Academic Press, 1979): 201-236.

[9] Scott Wolford, *The Politics of the First World War: A Course in Game Theory and International Security* (Cambridge: Cambridge University Press, 2019).

particular interest in the positive and negative externalities of US troop deployments overseas. His work has appeared in the *American Political Science Review*, *International Studies Quarterly*, *Journal of Conflict Resolution*, and *Foreign Policy Analysis*.

Bear F. Braumoeller is a Fellow of the American Association for the Advancement of Science and the holder of the Baronov and Timashev Chair in Data Analytics at The Ohio State University, where he serves as Professor in the Department of Political Science. His most recent book, *Only the Dead: The Persistence of War in the Modern Age* (Oxford, 2019), evaluates the empirical support for the decline-of-war thesis advocated by scholars like Steven Pinker and argues that patterns of international order are better predictors of trends in warfare than the spread of empathy, morality, and reason. His current research combines theoretical, statistical, and case-study methods to explore the relationship between international order and international conflict.

Hein (Henk) Goemans received his Doctorandus (MA) diploma in Law from the University of Amsterdam in 1996, an MA in International Theory and History from SAIS, the Johns Hopkins University in 1986, having completed his first year in Bologna. He received his PhD in Political Science at the University of Chicago, with a dissertation on War Termination, under the guidance of Steve Walt (Chair), John Mearsheimer, James Fearon and Duncan Snidal. His first two books focused on the role of leaders in decisions to terminate or start wars. His current long-term research project is on territorial conflict. Goemans uses mixed methods in his research.

Xiaoli Guo is an Assistant Professor at the Shanghai Institute of International Organization and Global Governance of the Shanghai University of Finance and Economics. She uses game theory, experiments, and quantitative methods to study authoritarian regimes, Chinese politics, and law and economics. In 2018-2019, she received the ISG Teaching Excellence Award, the highest honor the student body has for educators at UIUC, when she was a lecturer there.

Scott Wolford is Professor of Government at the University of Texas at Austin, Editor of *Conflict Management and Peace Science*, and Co-Director of the Correlates of War Project. His research focuses on international conflict, coalition politics, and game theory. His two books, *The Politics of Military Coalitions* (2015) and *The Politics of the First World War: A Course in Game Theory and International Security* (2019), are published by Cambridge University Press.

Essay by Bear F. Braumoeller, The Ohio State University

Teaching theoretical modeling in the social sciences is rare, and its rarity is odd. Theoretical models, like the Bohr model of the atom in physics or classical thermodynamics in chemistry, convey the fundamental domain knowledge of most disciplines. Plenty of theoretical models exist in history and the social sciences: one could list the Prisoner's Dilemma,[1] the security dilemma,[2] world-systems theory,[3] prospect theory,[4] the Klein-Goldberger model of the American economy,[5] the Daisyworld model of homeostasis,[6] and Granovetter's threshold models of behavior,[7] just to name a few. Such models are often quite influential in their respective disciplines, at least judging by the citation counts of works like these. The theoretical insights they convey are arguably their disciplines' core products.

Yet few social science departments systematically teach their students how to build new ones.

One would think that this state of affairs would be alarming, but it seems to occasion little concern. To be sure, there are occasional books about theorizing,[8] modeling,[9] and mechanisms.[10] But coursework on theory-building remains far less common than coursework in theory-testing, and sequences in theory-building are all but nonexistent.

There is more to this complaint than just the grumbling of a graying academic shaking his fist at a passing cloud. The paucity of instruction on theorizing reflects, I think, an orientation toward science that does not fully understand the value of theorizing, or even what theorizing is. In my experience, the pervasiveness of that orientation is the most challenging thing about teaching students to theorize.

* * *

In a 1998 essay, the sociologist Aage B. Sørensen argued that quantitative sociology had become more theory-poor over the previous three or four decades. He summed up the reasons for its decline with tart precision:

[1] Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984).

[2] John H. Herz, "Idealist Internationalism and the Security Dilemma," *World Politics*, 2:2 (1950): 157–80 https://doi.org/10.2307/2009187; Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976).

[3] Immanuel Wallerstein, *The Modern World System* (New York: Academic Press, 1974).

[4] Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* 47 (1979): 263–91.

[5] Lawrence R. Klein and Arthur S. Goldberger, *An Econometric Model for the United States, 1929-1952* (Amsterdam: North-Hollard, 1955).

[6] Andrew J. Watson and James E. Lovelock, "Biological Homeostasis of the Global Environment: The Parable of Daisyworld," *Tellus B*, 35B.4 (1983), 284–89 https://doi.org/10.1111/j.1600-0889.1983.tb00031.x.

[7] Mark Granovetter, "Threshold Models of Collective Behavior," *American Journal of Sociology*, 83:6 (1978): 1420–43 https://doi.org/10.1086/226707.

[8] James Jaccard and Jacob Jacoby, *Theory Construction and Model-Building Skills: A Practical Guide for Social Scientists*, Methodology in the Social Sciences (New York: Guilford Press, 2010); James N. Rosenau and Mary Durfee, *Thinking Theory Thoroughly: Coherent Approaches to an Incoherent World*, 2nd ed (Boulder, Colo: Westview Press, 2000).

[9] Kevin A. Clarke and David M. Primo, *A Model Discipline: Political Science and the Logic of Representations* (Oxford ; New York: Oxford University Press, 2012); *Models as Mediators: Perspectives on Natural and Social Sciences*, ed. by Mary S. Morgan and Margaret Morrison, Ideas in Context (Cambridge ; New York: Cambridge University Press, 1999); Scott E. Page, *The Model Thinker: What You Need to Know to Make Data Work for You*, First edition (New York: Basic Books, 2018).

[10] Jon Elster, *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences* (Cambridge ; New York: Cambridge University Press, 2007); *Social Mechanisms: An Analytical Approach to Social Theory*, ed. by Peter Hedström and Richard Swedberg (Cambridge: Cambridge University Press, 1998).

> The reason quantitative sociology has become theoretically poor is that the enormous progress in methodological power has turned quantitative methodology into a branch of statistics. This has led to a fascination, if not an obsession, with statistical models and concerns, and a neglect of the need to develop sociological models mirroring conceptions of mechanisms of social processes.[11]

To be clear, Sørensen was no opponent of statistical methods in the social sciences. (Neither am I.) Rather, he was frustrated with the extent to which statistical models have displaced theoretical models in his field. That is a complicated argument to articulate, because the distinction between statistical and theoretical models isn't as clean as one might like. But let me give it a shot.

Models, in the first instance, are mental representations of real-world processes or mechanisms. Their function is to allow us to understand something by simplifying it. Abstraction is, therefore, essential to modeling, because it is impossible to grasp the full details of even fairly straightforward processes.[12]

Even those who are on board with that statement, though, would most likely balk at one of its implications: Models must be wrong in order to be useful. They should not be wrong in important ways, but it is essential that they be wrong in unimportant ones. Beyond a certain level of spare realism, adding nuance hurts theory.[13] Even if it were attainable, perfect realism would be undesirable in a model. Jorge Luis Borges conveys this point memorably in his short parable, *On Exactitude in Science*,[14] in which cartographers in an ancient empire make a map that is as large as the empire—a map that is totally useless *because* it is perfectly accurate.

Abstract mental models, once formed, cannot be conveyed by telepathy. They have to be translated into a medium other than thought—generally, words, mathematics, a computational model, or a statistical one. Something is typically lost in translation, as it is between two natural languages. Sometimes it is translated into a verbal or mathematical model and then into a statistical model, which involves more slippage still. In the end, it is translated back into a mental model when others attempt to understand it.

It is important to emphasize that, despite the fact that models take all of these different forms, theoretical models and models of data are distinct kinds of models, each with its own logic. Ideally, they are compatible, like a hand and a glove, but that compatibility shouldn't obscure the fact that hands and gloves are different things. Statistical models do *reflect* theoretical logic, of a sort; they just aren't usually built with theory in mind. The foundational models in statistics—regression, logit, and the like—are designed to be models of distributions, not models of processes or mechanisms. They might be reasonable representations of processes or mechanisms; logit and probit, for example, work well as random utility models.[15] But if they are, it is generally by accident, and the representation is typically approximate. To extend the clothing analogy, they are the convertible wrap dresses of the modeling world—flexible enough to serve as a dress, a shawl, a scarf, a blouse, etc., but not as good as an item of clothing that had been tailored for one specific purpose.

---

[11] Aage B. Sørensen, "Theoretical Mechanisms and the Empirical Study of Social Processes," in *Social Mechanisms: An Analytical Approach to Social Theory*, ed. by Peter Hedström and Richard Swedberg (Cambridge: Cambridge University Press, 1998): 238–66 (238–39).

[12] Historians might roll their eyes at a social scientist calling for more abstraction, but let's face it—psychologists would look askance at the historians, neurobiologists would shake their heads sadly at the psychologists, and chemists would just throw up their hands at the lot of us, so let's just not do this.

[13] Kieran Healy, "Fuck Nuance," *Sociological Theory* 35:2: (2017):118–27.

[14] Jorge Luis Borges, "On Exactitude in Science," in *Collected Fictions* (New York: Viking, 1998).

[15] Christopher H. Achen, "Toward a New Political Methodology: Microfoundations and ART," *Annual Review of Political Science*, 5:1 (2002), 423–50 https://doi.org/10.1146/annurev.polisci.5.112801.080943.

Lacking concrete examples of more intentional theoretical models, econometrics students latch on to basic statistical models as if those models were designed with theory in mind—or worse, as if they were an adequate representation of *any* theory one might have. From there, theorizing reduces to justifying the inclusion of variables in a statistical equation. Once students have become "beta theorists" (in that all their theories take the form $y=X\beta$), they don't just fail to see the need for more theory—they develop instincts that are *antithetical* to theorizing, like adding more variables when fewer are needed,[16] or 'explaining' by accounting for as much detail as possible rather than by abstracting away from reality. This statistical mindset has become so pervasive that theories that don't neatly fit into a statistical mold, or that are designed with parsimony rather than explained variance in mind, have increasingly become foreign. That entrenched foreignness is a major barrier to teaching students to theorize.

Now, methodologists might protest, "What about more advanced statistical models, like selection models and structural equation models? Those are built with theory in mind!" Some more advanced statistical models do help, it is true, but others exacerbate the problem. The recent revolution in causal inference in the social sciences, for example,[17] while it has induced a healthy caution about the challenges of causal inference with observational data, brings with it a widespread and sometimes exclusive belief in a simple, monocausal manipulation account of causation.[18] The advent of machine learning techniques has prompted some enthusiasts to wonder whether we even need theory any more.[19]

While developments like these are unhelpful, the displacement of theory by statistical models predates them by quite a bit. Well before Sørensen penned his surgically precise takedown of statistical thinking in sociology (really, it's worth reading in its entirety), Edward A. Shils complained that

> the fact that the correlations among the indices of ambiguous analytical meaning is often high and that the possibilities of successful practical manipulation are thus enhanced constitutes a barrier to our perception of the need for theory.[20]

The fact that Shils wrote this *in 1949,* in the preface to Max Weber's *The Methodology of the Social Sciences,* might lead you to conclude that the social sciences are well and truly screwed when it comes to theorizing. It might prompt you to anticipate a diatribe about the death of knowledge, the feckless behavior of kids these days, the decline in good old-fashioned research, ah for the good old days, *fist shake, fist shake.* From my perspective, though, nothing could be further from the truth. Kids these days are incredibly smart, and the research they produce is fantastic. And overall, I'm more hopeful about theoretical modeling than I have ever been.

One of the biggest barriers to theoretical modeling has always been that it generally has to be done by hand. Social scientists in the 1970s could use computers to run regressions without having to do any math. By contrast, Thomas Schelling worked out the first agent-based model, his neighborhood segregation model,

---

[16] Christopher H. Achen, "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong," *Conflict Management and Peace Science*, 22:4 (2005): 327–39 https://doi.org/10.1080/07388940500339167.

[17] Luke Keele, "The Statistics of Causal Inference: A View from Political Methodology," *Political Analysis*, 23:3 (2015): 313–35 https://doi.org/10.1093/pan/mpv007.

[18] G. H. von Wright, *Explanation and Understanding*, Contemporary Philosophy (Ithaca, N.Y: Cornell University Press, 1971); Donald B. Rubin, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66:5 (1974): 688–701 https://doi.org/10.1037/h0037350. For a good critique, see Michael P. Keane, "Structural vs. Atheoretic Approaches to Econometrics." *Journal of Econometrics* 156:1 (2010): 3–20.

[19] Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, 23 June 2008 https://www.wired.com/2008/06/pb-theory/ [accessed 15 March 2022].

[20] Edward A. Shils, "Foreword." In Weber, Max, *The Methodology of the Social Sciences*, Glencoe, IL: Free Press, (1949): iii–x, at vii.

with pocket change and a grid drawn on a piece of paper.[21] While some standardization has occurred since then, theoretical models remain artisanal products—invariably small-batch and often lovingly handcrafted.

The kind of curiosity and creativity that led Schelling to experiment with pennies and dimes is hard to teach. In the interim, we have developed tools that are far more effective than pencils and pennies. Foremost among these is a free, open-source application called NetLogo,[22] which can be used to create agent-based models, computational network models, and systems dynamics models. Built around Logo, an educational programming language with simple syntax and informative error messages, the NetLogo program is fast and flexible enough to serve as an integrated development environment for theory and has a huge model library to serve as inspiration.

Coding might seem like an odd way to theorize, but it has many virtues. More than anything else, it is abstract, and therefore flexible. The agents in the model can be individuals, groups, cities, states, or whatever the theorist wants them to be. They have as few or as many attributes as the theorist wants them to have, and they can be programmed to interact conditional on those attributes. They can form ties to one another, and those ties can represent a wide range of relationships, from alliances to group membership to channels of communication. When they act, or interact, they follow rules that are as simple or as complex as the theorist makes them, and if they interact across some terrain, that terrain can have attributes as well.

What this flexibility implies is that the sorts of things that are often thought to drive individual and aggregate human behavior—desires, opportunities, beliefs, dispositions, situations, emotions, calculations, habits, and the like—can be represented in abstract form in simulated agents. Once the agents have been programmed, simulations can be run quickly, and the theorist can make adjustments based on the result. This sort of fast prototyping of theoretical models accelerates theorizing, and the ability to track results in real time helps the theorist spot interesting, unanticipated, or emergent behavior. (It is also *much* more fun to watch agents interact than it is to run a regression, and fun is a huge boon to pedagogy.)

Some things are easier to represent than others in computational models, of course. It can be difficult to program the sorts of complex decisions that are common in game-theoretic models, for example, and you don't get tidy, closed-form solutions that help you characterize the equilibrium behavior of a system. But because modern computers are very fast, you can typically brute-force your way to inference by running lots of simulations very quickly, or by changing one thing in the middle of a set of simulations and watching what happens, and in so doing you may find interesting things happening outside of equilibrium.

Finally, simulation encourages good theorizing, because the things that make for an understandable simulation also make for an understandable theory. You can add as many variables and decision rules as you want, for example, but when you do, you'll quickly understand how difficult even a simple model can be to interpret. In theoretical modeling, as in architecture, less is more.

I have run courses in computational modeling for nearly a decade now, and I have found it to be the fastest and most effective way to teach students to theorize. The learning curve is gentle, and the payoffs are easy to see. More than anything else, though, they learn theory-building as an activity that is entirely distinct from

[21] Thomas C. Schelling, "Dynamic Models of Segregation," *The Journal of Mathematical Sociology*, 1:2 (1971): 143–86; Thomas C. Schelling, *Micromotives and Macrobehavior* (New York: W. W. Norton, 1978): 147–50.

[22] Uri Wilensky, *NetLogo* (Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999) http://ccl.northwestern.edu/netlogo/; Uri Wilensky and William Rand, *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with Netlogo* (Cambridge, MA: The MIT Press, 2015).

theory-testing, with its own logic and *desiderata*. In the long run, that understanding will be far more important to them than their technical skills.

Essay by Hein Goemans, University of Rochester

Science is *organized* knowledge. But students, even graduate students, have little or no clue how to organize their knowledge. In their papers and in-class discussions, students grasp for arguments and insights that often are incoherent and contradictory. For at least two reasons, teaching formal theory from the ground up shows students a particularly powerful way to organize their knowledge and produce coherent and logically consistent arguments. First, formal theory clearly identifies the fundamental and essential assumptions upon which all formal theory is built and thus specifies what falls within its purview and what falls outside of it. Second, by starting from the very basics and systematically introduce refinements, student learn how knowledge can accumulate.

In the next section I first clarify what I see as the problem. I then proceed to lay out a case for the *usefulness* of formal theory to organize knowledge into a coherent worldview. It is important to acknowledge right off the bat that there of course exist other ways of organizing our knowledge of the social world, which often involve knowledge that is based on one inviolable authoritative text.

1 Is there a problem?

When I started my PhD at the University of Chicago, I certainly had no thought of organizing my knowledge, such as it was, or for the need of it. I had two master's degrees and read a lot about history and theories of international relations, but it all hung together like loose sand. I think it was in the beginning of my second semester that Daniel Verdier asked me what I hoped to accomplish in graduate school. I answered that I wanted to learn and study more theories of international relations. He told me I was wrong. Instead, he said, I was there to build my own worldview, my own coherent understanding of how the world worked. *He was right.* For most students, like me, building a coherent worldview typically proceeds through luck, haphazard accretion, and a large amount of time spent thinking (and not reading or preparing for exams). Much of this amounts to groping in the dark. I was lucky in that I had teachers, in particular John Mearsheimer, who had seemingly organized their own knowledge into an overarching framework. Students could adopt that or reject it. The latter in practice required that students *organize* their knowledge to develop consistent answers to the questions and challenges of the field and their fellow students.

True believers in 'paradigms' like Realism, Neo-Realism, Classical Realism, Neo-Classical Realism, Neo-Liberal Realism and Liberal Institutionalism would of course claim their preferred 'theory' offers an organized framework of knowledge. I don't want to re-hash any fruitless paradigm or intra-paradigm wars, which in the end amount to no more than glaring examples of the non-accumulation of knowledge. Instead I want to argue that such theories have undefined and pervious boundaries. To put it differently, from the supposedly same set of basic assumptions, different strands of theory derive different conclusions because they invoke (ad hoc) auxiliary assumptions.[1] But nothing in any of the theories sets any limit to the auxiliary assumptions that are allowed in or are to be excluded. This leaves students adrift.

1.1 The organization of formal models

We can do better, putting the organization of knowledge and the boundaries of our organizational framework up front. I therefore begin my intermediate class on "Mechanisms for International Relations" by starting from the foundations. Formal theory in IR is the study of strategic interaction of actors who by their joint actions produce outcomes.

[1] As Bob Powell showed again and again.

A transitional sentence here would be useful

- Actors are defined by their preferences and their beliefs.
- Preferences:

    1. For all outcomes, actors are able to say whether they prefer A to B, B to A, or are indifferent. In other words, their preferences are complete.
    2. Actors are not allowed to prefer outcome A to B, B to C and C to A; e.g., cycling is not allowed. In other words, preferences are transitive.

- Beliefs: what the actor thinks about the preferences of other actors.

The environment is the set of actions and information available to actors:

- Actions: physically describe what options are open to the actors.
- Information: what the actors know and what they can infer about others.

These are the foundations, building blocks, and boundaries of the rational choice program. These also make clear what falls outside the scope of this program. As I always tell my students, preferences *can* change, and sometimes aggregated (group) preferences are *in*-transitive. For the former, I tell students that I used to hate spinach, but now I love it. So my preferences have changed. For the latter, in Figure 1 I report polling in November 2018 of the UK's overall preference ordering. Take a good look: a majority of UK citizens preferred *Remain* over (former PM Theresa) *May's Deal*, *May's Deal* over *No Deal*, but also *No Deal* over *Remain*! These anecdotes show the boundaries of formal models, and what falls outside their scope. Other frameworks, with different boundaries, such as perhaps political psychology, could try to build knowledge around one or more violations of the rational choice approach. However, frameworks that contain internal contradictions are not (properly) organized.[2]

To organize your knowledge, you really shouldn't start in the middle and work backwards to first principles. Instead it is much easier and productive to start from first principles and build up your framework, brick-by-analytical-brick. The problem, as I see it, is that other putative organizing frameworks, such as 'Realism' are taught as if they spring forth fully formed from the head of Zeus. There are many problems with this approach, not least of which is the fact that no one knows whether or how to extend or build on such a fully formed 'theory.'

[2] What matters, as Robert Powell emphasized, is how *useful* these frameworks are for the question we want to ask and answer.

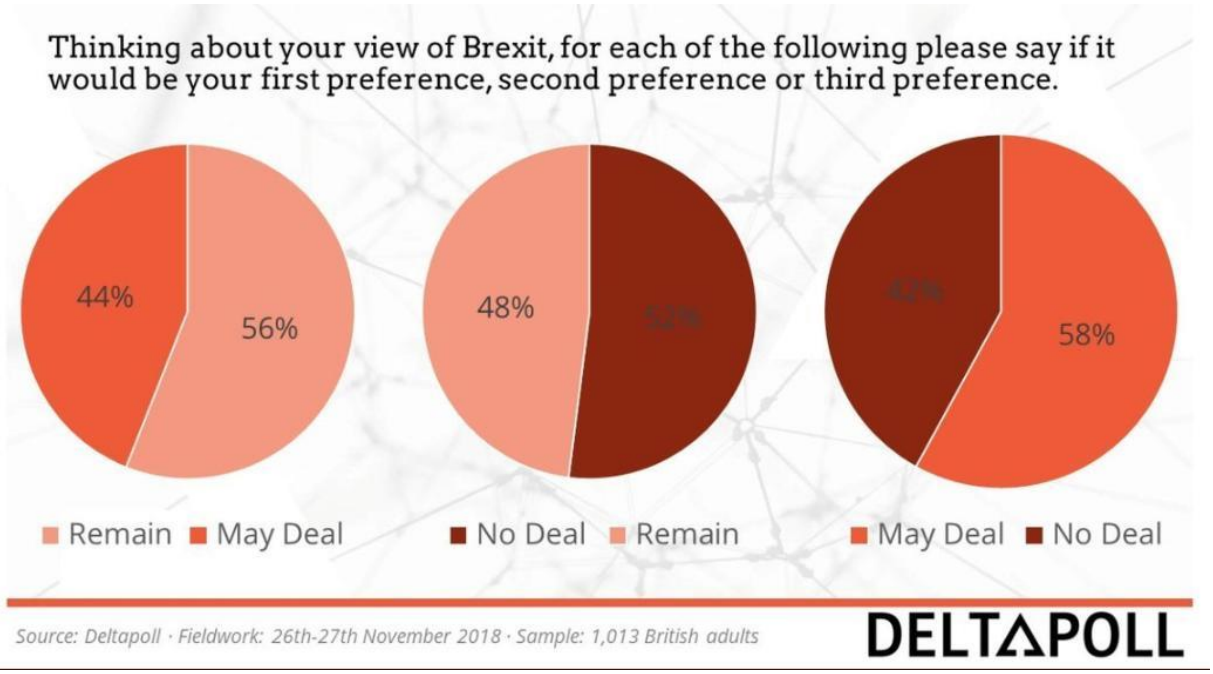

Figure 1: Brexit Polling

To illustrate the organizational power of building up from first principles, I rely on a classic game and a discussion of how I teach it. Let's look at the famous Prisoner's Dilemma (PD). The usual story is that two suspects in a major crime case are told by the prosecutor that there is enough evidence to convict each of them of a minor offense, but not enough evidence to convict either of them of the major crime, unless one of them acts as an informer against the other, in other words, unless one *Finks*. If both stay *Quiet*, each will be convicted of the minor offense and spend one year in prison. If one and only one of them finks, she will be freed, but must testify as a witness against the other, who will spend three years in prison. If they both fink, each will spend two years in prison. Let's now map these outcomes onto preferences. Clearly each prisoner's most preferred outcome is to walk free, and the worst outcome is to go to jail for three years. Thus we rank order preferences over outcomes as in Figure 2.

$$0 \textit{ years} \rightarrow 1^{st}$$

$$1 \textit{ year} \rightarrow 2^{nd}$$

$$2 \textit{ years} \rightarrow 3^{rd}$$

$$3 \textit{ years} \rightarrow 4^{th}$$

Figure 2: Mapping outcomes to preferences

Figure 3 presents the normal form game of the Prisoner's Dilemma. The surprising thing for students who see this for the first time is that although there is an outcome that *both* would prefer—e.g., their second best, one year in jail for each, where both stay quiet—they can't reach it, and the equilibrium outcome is their third best, two years in jail, where both fink.

Many teachers and the writers of the Wikipedia entry on the topic insist that it is necessary to tell the students that the suspects would be interviewed and confronted individually and separately. Obviously, or perhaps it is not so obvious, they are wrong. They argue that if the two prisoners sit in the same room and hear the same offer from the prosecutor, and are allowed even to communicate, they can agree to stay quiet. But please look at Figure 3 again. Why would anyone stay quiet? No matter what they *say*, each prisoner is still always better off finking. Let's say you and I are the two prisoners. You tell me we should both be quiet because then we'd both be better off! Let's say that I "believe" you. What will you do? If I stay quiet, you can choose to also stay quiet and go to jail for a year. Or you can fink and walk away scot-free, a choice you obviously prefer. But that would mean I would get my worst outcome, three years in jail. The problem here, in a nutshell, is that any promise to stay quiet is not enforceable. Allowing the prisoners to communicate and tell each other lots of stories changes *nothing*: not the choices or anyone's preferences. To this day, you will hear people say that the prisoners need to be separated, because somehow communication makes agreements enforceable. But to make agreements enforceable requires a very different game and that game decidedly is *not* the Prisoner's Dilemma. To build a theory from its foundations requires you to organize your knowledge. To start in the middle does not.

		Suspect 2	
		Quiet	*Fink*
Suspect 1	*Quiet*	$2^{nd}, 2^{nd}$	$4^{th}, 1^{st}$
	Fink	$1^{st}, 4^{th}$	$3^{rd}, 3^{rd}$

Figure 3: The Prisoner's Dilemma

To show how we can build onto the basic framework described above, let's design a game that does allow for enforcement, specifically, a game that allows a player to punish the other player if (s)he finks. To do so, we'll introduce extended form games. I first present the Prisoners' Dilemma in extended game form in Figure 4. As I tell the students, this *re*-organization introduces a *sequence* of choices, where one prisoner can choose first and the other chooses second.[3] We solve the game by backwards induction: starting from prisoner 2's choices at his last decision node (on either the left- or the right-hand side of the game tree). At either node, prisoner 2 is best off by choosing to *Fink*. Now moving back to prisoner 1's choice, that boils down to getting his 4th most preferred (worst) or his 3*rd* most preferred outcome, and he will naturally choose the latter. Again, the equilibrium outcome is *(Fink, Fink*) where each prisoner gets his third most preferred (second worst) outcome.

I amend the story as follows. The prisoners tell each other that if one of them *Finks*, and the other stays *Quiet*, the prisoner who stayed quiet has the opportunity to tell the prosecutor the details of another crime his partner committed. (The prosecutor confirms this.) Given that the prosecutor now has solved one additional case, he rewards the prisoner who *Punished* his fellow prisoner for *Finking* with one year off his sentence. The

[3] At this point, I ask them if they can give me examples where moving first or second gives one player an advantage. My favorite example of the advantage of moving last is "The Price is Right."

other prisoner who the prosecutor now can convict on two crimes gets five years. Let's map these outcomes into preferences in Figure 5.

We can now analyze the amended story, where one prisoner can punish the other for *Finking* and thereby enforce silence. Figure 6 presents *this* game.

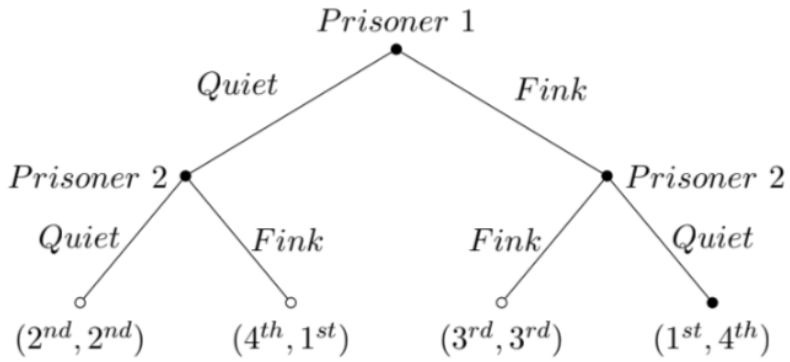

Figure 4: Prisoners's Dilemma in extended form

$$
\begin{array}{lcl}
0\ years & \rightarrow & 1^{st} \\
1\ year & \rightarrow & 2^{nd} \\
2\ years & \rightarrow & 3^{rd} \\
3\ years & \rightarrow & 4^{th} \\
5\ years & \rightarrow & 5^{th}
\end{array}
$$

Figure 5: Mapping outcomes to preferences - 2

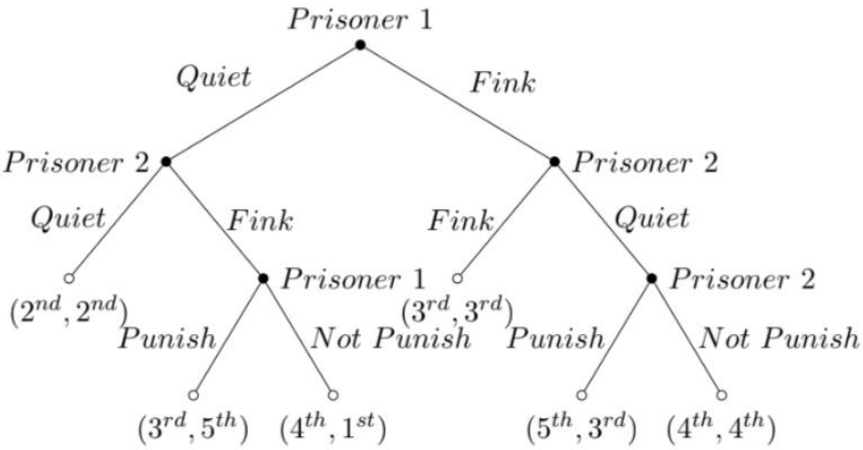

Figure 6: Prisoners with punishment

Starting from the right-hand side of the game tree, where Prisoner 1 has *Finked*, and Prisoner 2 chose to stay *Quiet*, we see that at that point, Prisoner 2 would prefer to *Punish*. Moving up to the choice to *Fink* or stay

Quiet we see that Prisoner 2 is indifferent between these two choices; in either case he gets his 3rd most preferred outcome. Going to the left-hand side of the game tree, if Prisoner 1 stays *Quiet* and Prisoner 2 chooses to *Fink*, Prisoner 1 would prefer to punish (he gets his 3rd worst rather than his 4th worst outcome). Given that choice by Prisoner 1, Prisoner 2 would prefer to stay *Quiet*. Now Prisoner 1 can compare the consequences of his choice to *Fink* or stay *Quiet*. If he *Finks*, the best he can get is his 3rd worst outcome. But if he stays *Quiet* he can get his second best outcome (and so does Prisoner 2!). Note the following:

- The story had to be changed. Each prisoner now has to know about another crime his partner committed. In addition, the prisoner who stayed *Quiet* after the other guy *Finked* gets another turn to speak to the prosecutor.
- This is a very different game than the original PD in Figure 3. We now have an extensive form game and the Subgame Perfect equilibrium concept as compared to the original normal form game and Nash Equilibrium.

In other words, the organization of the story is different, and to match it, the organization of the game has to change as well. Once we re-organize the original PD to include the assertion that the ability to communicate is essential, we obtain a radically different situation. At least two new findings emerge:

- Prisoner 2 is now indifferent between Finking and staying *Quiet* after Prisoner 1 *Finks*.
- Both Prisoners are better off by introducing the possibility of punishment for *Finking*.

Our revised scenario is a *lot* more complicated that the original Prisoner's Dilemma presented in Figure 3. To get to the point where prisoners can credibly enforce any promises to stay quiet requires additional machinery: sequential choice, the Subgame Perfect Equilibrium solution concept, and a modicum of creativity in turning the story into a game.

This example shows students that formal theory can help them learn how to organize their knowledge. First, it shows that the claim that the prisoners cannot communicate is an unnecessary—and even confusing—element of the story used to illustrate the Prisoner's Dilemma. It then shows us how we can the tools of formal theory to show when that element is necessary and changes the outcome. It show how we can build on, extend, and modify the game using the basic assumptions we identified before and how these things hang together. It thus shows what is essential and what is not.

2 Conclusion

This brief essay describes two features of formal theory that highlight the fundamental insight that science is organized knowledge. It demonstrates how formal theory sets clear boundaries of its scope and how starting from first principles allows us to organize knowledge. Formal theory is a useful example of a way to organize one's knowledge, but it is by no means the only one.

Essay by Xiaoli Guo, Shanghai University of Finance and Economics

The game theory course "Strategic Models of Political Science" I offer for undergraduates covers material related to international relations such as the international arms race, crisis bargaining, and collective action between countries, as well as other topics such as voting and elections, the provision of public goods, and so on. For me, there is no essential difference in using game theory to teach international relations and other topics. Game theory emphasizes thinking in a rational way, assuming that all participants are strategic, and considering the interactions of all participants in the decisionmaking process. Therefore, the perspective of game theory applies not only to how to treat friends or opponents, how to decide which candidate to vote for, but also to how to observe and think about cooperation and conflict between countries. Whether or not the students taking my course will study or work in the field of international relations in the future, I hope that their interest in this field will become a habit. To this end, I try to convey to the students the idea that 'international relations' is not an unreachable grand topic. On the contrary, we can understand it in the way of understanding things around us.

When applying game theory to international relations in undergraduate teaching, I often begin with scenarios familiar to the students, then connect the knowledge points behind them with political phenomena, and further go deeper into a more theoretical and abstract level. In this way, I provide a knowledge network in which students can integrate their own knowledge and establish links between various parts. I often employ more than one teaching method in one course, such as multimedia teaching, student presentations, in-class experiments, group work, etc. I also urge students to answer and raise questions, and engage in a rolling discussion, which step-by-step leads to the knowledge point I intend to highlight. The Q&A style of teaching engages students' attention, enhances their confidence in mastering the materials, and reveals their strength and weakness in learning, to which I can refer in the future and make adjustments. In my experience, students learn actively when they have a useful network of information to build upon, when the information is presented repeatedly in various formats, and when they can engage in walking through the logic, drawing the conclusions, and applying them to real world scenarios.

Let me take my experience in teaching the most famous game, the Prisoner's Dilemma, as an example. After introducing the features and the equilibrium of the game, I show the video *Golden Balls*[1] to the students. In the video, two players have a chance to win £13,600—if both choose Split, each one receives half the money; if both choose Steal, neither player wins any money; and if one player chooses Steal and the other player chooses Split, the Steal player wins all and the Split player gets nothing. This game is a variant of the Prisoner's Dilemma. When the students watched the video, they usually enjoyed it, and the unexpected ending often caused exclamation and laughter. At the end of the video, I would tell the students to formalize the game, put the payoffs into a matrix, and see what the equilibrium is. I would ask the students to think and answer whether the game we constructed was a Prisoner's Dilemma. If not, what additional assumptions should we make to turn it into a Prisoner's Dilemma? Was the outcome of the *Golden Balls* game we saw the same as the equilibrium of the Prisoner's Dilemma? If not, then why? These questions usually provoked much discussion. The students were often excited that some basic game theory knowledge we just learned could be used to explain real and interesting problems in life.

At that point, I turn to international relations and tell the students that this simple and classic game can explain cooperation and conflict between countries as well. I introduce the students to the long-term arms race between the North Atlantic Treaty Organization led by the United States and the Warsaw Treaty Organization led by the Soviet Union from the end of World War II to the disintegration of the Soviet Union. The international arms race is undoubtedly much more complicated than the "Golden Balls" game, but the logic behind the two has something in common. Similarly, we build a game about the US-Soviet arms race, discussing the payoff structures of the two countries and under what circumstances the game was a

[1] https://www.youtube.com/watch?v=S0qjK3TWZE8&t=16s. Accessed July 25, 2022.

Prisoner's Dilemma. To that end, I tell the students more about that history, especially the costs of the arms race between the two countries. For example, there were differences in industrial structures between the capitalist United States and the Communist Soviet Union at that time. The Soviet Union's adherence to the heavy industry-centric strategy put enormous pressure on the daily lives of its people, and the United States had a stronger economy to withstand that pressure.

I ask the students to think about how the economies and the stability of the two countries would be affected if the game were repeated, that is, if the arms race continued. Some students would point out that the arms race was closely related to the disintegration of the Soviet Union. I ask them why, if the Soviet Union was rational, it did not choose to reduce its military expenditures. Through these discussions, the students would gain a deeper understanding of why the game reflects a "dilemma"—why the players could not just both pick Split, or why the two superpowers could not stop the arms race and make better use of their money. I would further point out that the Prisoner's Dilemma deeply captures the conflict between individual rationality and collective rationality—all players are rational, but they cannot achieve a result that maximizes the interests of both sides.

For most topics for my game theory course, I have designed corresponding games, videos, and group work. For example, for the Matching Pennies game,[2] I will prepare some Chinese coins as props and prizes and ask several pairs of students to play the game in front of their classmates. For the Battle of the Sexes game,[3] I will have several pairs of students play it repeatedly. Sometimes, they can develop mutually beneficial and fair cooperation. Sometimes more aggressive students insist on choosing an action to force the other players to coordinate with them. I often interview students and ask them what strategies they used when playing the game. Having students see different types of players and different strategies can be very helpful for them to learn the materials with a critical mind.

Similarly, I encourage students to look at the topics of international relations related to these games from different perspectives, whether they are crisis negotiations, nuclear competition, collective action between countries, or the role of international organizations, etc. I tell the students that different schools of international relations theory have different explanations for cooperation and conflict between countries based on different assumptions. For example, Realism emphasizes the role of power on state behavior, focusing on the power balance between nation states and the pursuit of the interests of the state.[4] Liberalism, on the contrary, denies the absolute role of power in international relations, and emphasizes mutual benefits and international cooperation, with the assistance of international organizations and nongovernmental actors.[5] There is no absolute right or wrong in these views. Sometimes the Prisoner's Dilemma can well reflect the dilemma faced by international cooperation. Sometimes, however, the Stag Hunt game[6] can better capture the fact the countries can achieve a win-win situation through cooperation and that the difficulty lies

[2] In the Matching Pennies game, each of two players has a penny and secretly turns the penny to heads or tails. The players then simultaneously reveal their choices. If the pennies match, i.e., two heads or two tails, player 1 wins and keeps both pennies; otherwise, player 2 wins and keeps both pennies. This game has no pure-strategy Nash equilibrium.

[3] In the Battle of the Sexes game, a man and a woman choose between two events to attend, a football game and a ballet. The man prefers the football game while the woman prefers the ballet. Yet, both of them believe that the value of accompanying each other is higher than the event itself. The game has two pure-strategy Nash equilibria—both attend the football game or both attend the ballet.

[4] For a comprehensive review of the realism school, see Jack Donnelly, *Realism and International Relations*. (New York: Cambridge University Press, 2000).

[5] Eric Shiraev and Vladislav M. Zubok, *International Relations*. (New York: Oxford University Press, 2015): 77-112. Also see Andrew Moravcsik, "The New Liberalism," in *The Oxford Handbook of Political Science*. (New York: Oxford University Press, 2011): 709–729.

[6] In the Stag Hunt game, two hunters decide independently whether to hunt a stag or a hare. One hunter can catch a hare on his own, which is less valuable than half a stag. To successfully hunt a stag, the two hunters must cooperate and help each other. There are two pure-strategy Nash equilibria of the game—both cooperate to hunt the stag, or each hunter relies on himself to catch the hare.

in their lack of mutual trust. I teach international relations from multiple perspectives and often keep open conclusions, because I want the students to be open-minded and critical.

In addition, I often tell the students that game theory emphasizes rationality and utility-maximization, but in reality, behaviors deviating those assumptions are often observed. For example, in the Ultimatum game,[7] a rational proposer should offer the smallest amount of money, say one penny, to the responder, and the responder should accept it—it is better than nothing. However, numerous experiments have demonstrated that responders value being treated fairly and prefer to reject an excessively low offer.[8] Similarly, in experiments of the Dictator game,[9] allocators display high level of altruism that is not predicted by the Nash equilibrium.[10] One main lesson that I hope my students learn is that besides rationality, we need to take into account the players' sense of fairness, altruism, willpower, biases and other factors, both in daily life and in our thinking about international relations. Thus, learning to interact rationally, although important, to maintain a positive and cooperative mindset is often beneficial.

[7] There are two players in the Ultimatum game, a proposer and a responder. The proposer is endowed with an amount of money, and makes an offer to the responder. The responder can accept or reject. If the offer is accepted, the money is split per the division. If the offer is rejected, both players receive nothing.

[8] See, e.g., Werner Güth, Rolf Schmittberger, and Bernd Schwarze. "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior & Organization* 3:4 (1982): 367-388, DOI: https://doi.org/10.1016/0167-2681(82)90011-7; Hessel Oosterbeek, Randolph Sloof, and Gijs Van De Kuilen. "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-analysis," *Experimental Economics* 7:2 (2004): 171-188, DOI: https://doi.org/10.1023/B:EXEC.0000026978.14316.74.

[9] In the Dictator game, one player (the allocator) receives an endowment and then makes an offer, which can be zero, to another player (the recipient), who cannot choose to reject.

[10] See, e.g., Robert Forsythe, et al. "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior* 6:3 (1994): 347-369, DOI: https://doi.org/10.1006/game.1994.1021; Gary E. Bolton, Elena Katok, and Rami Zwick. "Dictator Game Giving: Rules of Fairness versus Acts of Kindness," *International Journal of Game Theory* 27:2 (1998): 269-299, DOI: https://doi.org/10.1007/s001820050072.

Essay by Scott Wolford, The University of Texas at Austin

A few years ago, I decided to bridge the way I research International Relations with how I teach international relations—which, for me, meant international conflict and security. That mismatch was frustrating, because I want undergraduates to learn how to do political science, to know why political science says something, not just what political science says about a particular topic. That is not always easy, as evidenced by the fact that we spend so much time training PhD students in research methods. Nonetheless, I wanted my students to be better at evaluating explanations and making their own. I wanted them to learn how to consume game-theoretic work and to write and analyze their own games, so they'd have better skills in the logic of inquiry and explanation and so they'd have a leg up if they ended up pursuing a PhD. That meant I needed to stop simply translating the insights of formal theory in class, which never really let students see under the hood. Rather, I needed to teach a full undergraduate formal theory course alongside a full undergraduate course on international security.[1] So, I set about tearing down my international security course and rebuilding it into something completely different.[2]

The resulting course, "World War I in Real Time," used the single, unifying narrative of the eponymous war to (a) generate a series of puzzles from a narrative history of the conflict that (b) we solved, actively and in class, with the tools of game theory. Using the Great War had a lot to do with the then-approaching centennial of its outbreak, but the war is great as a teaching tool because (a) students typically don't know much about it coming in and (b) it has almost everything when it comes to important concepts and processes in international relations.[3] My initial rationale for the course was that students would read only history, identify interesting puzzles where events or decisions cut against our intuition or received wisdom. Then, we would do the science, that is, specifying, solving, and interpreting games, together in class. This ensured that (a) the tools of formal theory would be introduced with a clear justification for and demonstration of their use, (b) students would learn something about both the "urkatastrophe of the 20th century"[4] and (c) the accumulated knowledge of IR research, and (d) they would have some practice writing down and solving their own games. In this essay, I summarize the rationale behind the course and its associated textbook.[5] And I hope to make the case that the substantive focus on one big, overarching narrative both keeps students engaged and makes the technical aspects of teaching game theory easier than they might otherwise be.

A typical week in the course involves the identification of a puzzle generated from reading narrative histories of the war. For example, why did all the major belligerents settle into strategies of attrition, which seemed to preclude any kind of military decision, leading to accusations that the generals were unimaginative butchers ordering frontline solders into a futile war? To solve the puzzle, we wrote down a game that took seriously the strategic tradeoffs between attrition and maneuver, showing that attrition was an equilibrium because it reduced casualties relative to the unrestrained (and shockingly costly) attempts at maneuver early in the war. And our focus on equilibrium as a causal standard highlighted that neither side wanted a return to the first months of fighting that wiped out almost entire peacetime armies. Attrition, which on the Western Front meant trench warfare, with all its horror and apparent senselessness, was an equilibrium until one side could benefit sufficiently from switching back to maneuver: in this case, the Allies, with updated tactics and a steady flow of American dollars, dreadnoughts, and doughboys after 1917. By putting ourselves into the heads of

[1] I promise it's not as quixotic as is sounds.

[2] https://youtu.be/0hYZaqYCZyQ

[3] No nuclear weapons, of course, but that's an easy topic to add as needed. Still, I'd be remiss not to note Cullen Hendrix's injunction about spending too much time with charismatic megafauna—even if this one makes us confront the idea of empire more than most WWI treatments like to recognize. See https://politicalviolenceataglance.org/2017/11/21/charismatic-megafauna-in-conflict-studies-or-why-wwii-is-the-giant-panda-of-the-conflict-security-field/

[4] https://adamtooze.substack.com/p/chartbook-57-1914-the-urkatastrophe

[5] Scott Wolford. *The Politics of the First World War: A Course in Game Theory and International Security*. (Cambridge: Cambridge University Press, 2019).

the generals and politicians who settled into the war's (long) attritional phase, students could understand one of the war's key puzzles (its operational futility), and they could see the value of equilibrium reasoning as a standard of explanation.[6]

After using a purpose-built game to solve the puzzle, I spend time showing how the puzzle and its answer relates to the international security literature, making the nomothetic case that events in the war can be understood as a specific instance of something broader, something more general, in international relations. So, the attrition week spends time talking about the link between military strategy and war duration. Other lectures/chapters relate their core games to theories of bargaining and war (both outbreak and duration), coordination and limits on fighting, the politics of empire, domestic politics and international cooperation, cooperation in alliances and military coalitions, even the ideational foundations of international orders. But the war and the literature are connected by the game, the representation of a strategic problem that helps us understand why the actors in our story chose as they did.

This focus on "real-time" explanation is the course's through-line. It forces us to abandon hindsight, to think about what the characters in our story knew and believed when they made their decisions. We have no truck with the facile 'it must've been a mistake, so let's hold these people to an informational standard that only we in the future have' approach. Game theory helps us think about the incentives and beliefs and options of the relevant players, and to humanize them and, yes, to rationalize their choices. Hindsight is excellent for identifying puzzles, but it is trash for solving them. To solve puzzles about IR, we need to think hard about what the actors saw, knew, and believed about their options and about each other. We must get in their heads, and game theory is a useful tool for doing just that. When we introduce new theoretical tools—like Bayesian Nash Equilibrium—it is clear from a substantive standpoint why we do so. My goal is to deter that one student from asking "Why do we need to learn this?" by identifying the need for new tools before we introduce them. For example, Subgame Perfect Equilibrium emerges as a way to think about incredible commitments to peace agreements, like Germany's incentives to defect from an armistice. Likewise, Bayesian games emerge as a way to think about Britain's decision to join the war after the invasion of Belgium, which altered public and elite beliefs about German war aims.

That is the positive case for the course, which tracks the increasing complexity of the war with increasing sophistication of the tools we use to represent it. But when the textbook underwent peer review,[7] I confronted some reasonable concerns about the approach, most of which revolved around the challenges of getting political science undergrads to care about 'doing math' or, more generally, the technical nature of the material. The course was made better for these reviews, one of which I will recount here.

A reviewer wrote, and this is almost verbatim, 'Students just don't care what a separating equilibrium is.' I interpreted this as a concern that technical material is a distraction, that students will be put off by the formalities of, in this case, Perfect Bayesian Equilibrium (PBE). All students really want to know, the story goes, is what the invasion of Belgium had to do with Britain joining the war. My rejoinder was that, sure, students may not care about separating equilibria before taking the class, or even before they get to the chapter about British entry. But the key is to show them why they're learning the formalities, to reveal these ostensibly technical tools as solutions to clear substantive problems; and to emphasize that they are learning these tools not for the math *per se*, but for the logic. The PBE issue makes this case stark, in fact: arguments about uncertainty and signaling that do not have a good approach to equilibrium reasoning—to thinking about how senders and receivers think about signals both sent and unsent—also struggle to produce

[6] In a related analysis, they could also see why politicians who were accountable to voters tried to shift the blame for attrition, and even the outbreak of war itself, onto the generals after the war.

[7] There were twelve reviewers. *Twelve*!

satisfying explanations. So, the hope is that the course convinces students to care about what a separating equilibrium is.

When students know that we need a certain tool for a certain reason, for example, to represent something essential to a particular puzzle (in this case, uncertainty), then learning how to use that new tool is no longer about the 'math.' It is about how we represent players' uncertainty without doing too much violence to what we wish to represent and without smuggling distortionary assumptions or sloppy reasoning into our story. When I introduce a puzzle and uncover a specific need—how do we model uncertainty and changing beliefs in a useful way?—the formalities can be seen for what they are: secondary to the substantive, explanatory uses to which we put them. I bring this up not to criticize that reviewer (it was a reasonable concern) but to relay a common view of the challenges of teaching game theory. My solution is to dive deeper into the historical record, to think harder about the building blocks of theory and explanation as we use them. The key, at least for me, is to show students two boards that need to be connected before I show them a hammer and a nail. Introduced in their proper substantive context, apparent arcana like types of Perfect Bayesian Equilibrium emerge as critical tools in the act of representation, of understanding, and of learning about international relations.

I want to conclude with one caveat and one bit of encouraging feedback about the course. First, I am aware that calling a course "World War I in Real Time" has a particular effect on who's likely to sign up for it. It is a course *I* would have taken as an undergraduate: designed for war geeks (for lack of a better term), who are alternately fascinated and horrified by war, but not implying any kind of formal theory component.[8] I've done it for years, though, and the course is still (despite my excess enthusiasm for the material, low-brow *Star Wars* jokes, and unrelenting "30 Rock" references[9]) popular; it has also sent a few students to grad school in political science (or so they tell me). Second, the last time I taught "World War I in Real Time" was Spring 2020, when we moved midway through the semester to online teaching. The move was easy, and my employer was great at helping us all make the transition, but the modal comment in end-of-semester evaluations entailed disappointment that I had had to cut back on the formal, mathematical elements of the course. That, to me, was telling: formal theory and IR can be taught together, and one can strengthen the teaching of the other, with an overarching narrative like the First World War.

I can't wait to do it again.

[8] I study war because I fear it, full stop.
[9] https://youtu.be/04a3bnpV9eY