The Behavioral Revolution and International Relations

Emilie M. Hafner-Burton, Stephan Haggard, David A. Lake, and David G. Victor

What explains the strategically costly and ill-planned American invasion and occupation of Iraq? What accounts for Saddam Hussein's failure to take actions that might have deflected it? These decisions can be explored with rationalist tools, including the existence of credible commitment problems and asymmetries in information.[1] But explanations of this sort beg a number of important questions. The Clinton and Bush administrations did not differ substantially in their information about Iraq. But Bush administration officials—and the president himself—did hold *beliefs* that differed substantially from those of their predecessors, and those beliefs had profound effects.

Decision making by both Iraqi and US leaders displayed strong biases. Saddam Hussein failed to recognize that the United States was committed to war unless he was willing to reveal credibly that he had, in fact, dismantled his weapons of mass destruction. The United States signaled its intentions repeatedly, but the Iraqi leader remained impervious to new information. Bush administration officials believed that the Americans would be greeted as liberators and democracy would flourish of its own accord. Such motivated reasoning both precipitated war and contributed to the failure to plan adequately for rebuilding the Iraqi state in war's wake. The causes of the Iraq War and the disastrous consequences of its aftermath appear to lie as much in the realm of beliefs and decision making as in standard theories of bargaining.

Similar anomalies can be found in the study of international political economy. The theory of open-economy politics offers clear predictions about individual preferences with respect to trade policy. When factors of production are specific to an industry,

We thank Rick Hermann, Michael Horowitz, Josh Kertzer, Andy Kydd, Rose McDermott, Brad LeVeck, Jon Pevehouse, Brian Rathbun, Jonathan Renshon, Elizabeth Saunders, Phillip Tetlock, Dustin Tingley, Michael Tomz, and Keren Yarhi-Milo for detailed and helpful comments and background discussions, Robert Powell for detailed commentary on the project and underlying papers, and Steve Carlson, Alex Hughes, Shannon Carcelli, Brandon Merrell, and Linda Wong for their invaluable research assistance. We are indebted to the Laboratory on International Law and Regulation, the School of Global Policy and Strategy, the UC Institute on Global Conflict and Cooperation, the Norwegian Research Foundation, and the Jerri-Ann and Gary E. Jacobs Chair in Social Sciences for financial support for this project.

1. Debs and Monteiro 2014 and Lake 2010–11 provide a dynamic model of asymmetric information. For an alternative account of decision making leading up to the war, see Harvey 2012.

International Organization 71, Supplement 2017, pp. S1–S31
© The IO Foundation, 2017.
doi:10.1017/S0020818316000400

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

individuals employed or invested in the comparatively advantaged sector should favor free trade. When factors of production are mobile, individuals who possess assets that are relatively scarce should favor protectionism.[2] Yet, after more than a decade of careful empirical research, there is little evidence that voters actually define their interests in these rational, materialist ways. Instead, citizens' preferences appear to be in part sociotropic—rooted in concerns about the economy's performance as a whole and averse to policies that harm the least advantaged in society.[3] Individuals also have predispositions with respect to trade that are rooted in nationalism, ethnocentrism, and even racism.[4] Women are more protectionist than men, perhaps because they are more averse to social inequalities.[5] Trade policy preferences depend heavily on how the issues are framed, for example, whether questions engage respondents' status as a producer or consumer.[6] Rather than holding attitudes determined by their position in the international market, individuals appear to be guided by dispositions rooted in emotion, social psychology, and even genetic differences.

Armed with these sorts of insights, a new behavioral revolution has swept across the social sciences in the last few decades.[7] With origins in psychology, of course, psychological models have fueled the dramatic growth of behavioral economics and are now gaining traction in political science as well.[8] The defining characteristic of this revolution has been the use of empirical research on preferences, beliefs, and decision making to modify choice- and game-theoretic models.[9]

This is hardly the first time that international relations scholars have looked at how decision making might affect political outcomes. Earlier literatures drawing on psychology took advantage of prospect theory and research on decision-making heuristics.[10] What is new in today's behavioral revolution is the explosion of experimental research in both laboratory and field settings. This empirical work has spawned important theoretical advances, such as a growing consensus around a "two-level" model of cognition in which some choices are intuitive and immediate—what Kahneman calls "System 1" or "fast" cognition—while others are slow, deliberative,

2. Frieden and Rogowski 1996; Hiscox 2002.
3. Lu, Scheve, and Slaughter 2012; Mansfield and Mutz 2009.
4. Guisinger 2014; Mansfield and Mutz 2009; Sabet 2014.
5. Burgoon and Hiscox 2004; Goldstein, Margalit, and Rivers 2008.
6. On framing effects, see Hiscox 2006; Naoi and Kume 2011; but also Ardanaz, Murillo, and Pinto 2013.
7. The first behavioral revolution arrived in the 1950s and 1960s when research shifted from the study of formal-legal institutions to individual and group action. Key scholars leading this first revolution were Philip Converse, Robert Dahl, David Easton, and Heinz Eulau, among many others. The current cognitive revolution begins with the seminal work of Kahneman and Tversky.
8. Camerer, Loewenstein, and Prelec 2005; DellaVigna 2009; Lopez, McDermott, and Petersen 2011.
9. We follow Della Vigna (2009) in structuring the behavioral economics literature around these three categories.
10. Prominent examples include Allison 1971; Janis 1972; Jervis 1970, 1976; Larson 1989; Levy 1997; McDermott 1998; Steinbruner 1977. Related work includes scholarship on the use and abuse of historical analogies and metaphors: Khong 1992; Neustadt and May 1988.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

and more "rational"—referred to as "System 2" or "slow" thinking.[11] Also new are the efforts to anchor behavioral observations in a stronger neurological foundation through advances in brain science.[12] The result has been a wealth of research presenting both complements and alternatives to rational choice models.

This special issue aims to chart a strategy for incorporating the new behavioral revolution more fully into the study of international relations.[13] In principle, standard rationalist approaches to the study of international relations have allowed individual preferences, beliefs, and decision making to vary.[14] In practice, however, the focus on structure and interstate games mitigated the need to explore these variations in any great detail.[15]

Although the behavioral revolution had its origin in debates over expected utility theory and rational choice, the insights from this revolution also open important opportunities for dialogue with constructivists. Constructivists have long questioned rationalist conceptions of how preferences are formed, posing an alternative model that emphasizes the social nature of preferences and processes of socialization.[16] Behavioralists and constructivists also both look to prevailing ideas, norms, heuristics, and logics of appropriateness as determinants of individual and social choice processes.

How broad of a challenge does the behavioral revolution pose to rationalist models? For some, the new behavioral revolution is a complement to existing models. Improved, empirically grounded assumptions about core parameters such as social and time preferences, propensities for risk, and beliefs about the world can be incorporated relatively easily into some existing choice- and game-theoretic models, although with important changes in anticipated equilibrium outcomes.[17] For example, some strands of behavioral research suggest conditions under which policy-makers and publics may be more accepting of risk, and thus more prone to engage in risky behaviors that could lead to conflict and war. Other strands of research help to explain why policy-makers and publics are often more cooperative and trusting in settings such as prisoners' dilemma games where defection would otherwise constitute an equilibrium strategy.[18] These findings extend the field's

11. Kahneman 2011. See also Kahneman and Tversky 1979; Tversky and Kahneman 1974. See also Haidt 2012 who uses the metaphor of a rider and an elephant to communicate the same basic point.

12. For example, Bechara, Damasio, and Damasio 2000; and Fehr and Camerer 2007.

13. For behavioral research in the field of international law, see Broude 2015; Galbraith 2013; Jolls, Sunstein, and Thaler 1998; Poulsen 2013; Sitaraman and Zionts 2015; van Aaken 2014.

14. For example, see research on leaders by Horowitz, McDermott, and Stam 2005 and on time horizons by Edelstein 2002 and Krebs and Rapport 2012.

15. Some notable exceptions include Horowitz, Stam, and Ellis 2015 and McDermott 2007.

16. For the contrast in rationalist and constructivist approaches to preferences, see Fearon and Wendt 2002. The central themes of this agenda were first spelled out in a cluster of both theoretical and empirical work appearing from the mid-1990s and covering both security and international political economy themes: Biersteker and Weber 1996; Checkel 1998; Finnemore 1996; Finnemore and Sikkink 1998; Hall 1999; Katzenstein 1996; Klotz 1996; Kratochwil 1989; McNamara 1999; and Wendt 1999.

17. Camerer 2003.

18. Hafner-Burton, Hughes, and Victor 2013.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

canonical models to help explain puzzles as wide ranging as confrontational behavior by weak powers to the noticeably high levels of compliance with rules lacking in credible enforcement mechanisms.[19]

For other scholars, the new behavioral revolution portends a much more fundamental challenge to rationalist models. Systematic failures in perception can fundamentally change the strategic setting or generate outcomes that are radically different from models rooted in the expectation of fully rational behavior. Framing, anchoring, and menu effects can influence choice as well as perceptions of adversaries, suggesting that analytical attention should focus on how decision-making processes are structured in the first place, both at the individual and group level. The influence of emotion on behavior may lead some actors even to violate core tenets of rational choice altogether, such as the transitivity of preferences. Such factors may affect political behavior by raising levels of uncertainty, impeding actors' ability to reliably signal and interpret signals of intentions, and introducing more stochastic or even purely random elements into strategic interactions.[20]

The research program we outline in this special issue does not seek to provide a unified framework or universal set of propositions deduced from a single overarching behavioral theory. No such general theory exists, even in cognate disciplines. What we can offer is a framework for integrating the new behavioral revolution into international relations. In doing so, two significant themes emerge. The first is the gains from focusing on the heterogeneity of preferences, beliefs, and decision-making processes *even across similarly situated individuals.* Much of the early research in behavioral economics focused on cognitive traits and biases that were widely distributed among subjects in experimental settings; the objective was to show the pervasive nature of bias. As the behavioral revolution is unfolding in international relations, however, empirical research is now focused on the causes and consequences of heterogeneity across relevant actors, including the extent to which behavioral traits converge or diverge from rationalist expectations.[21] The range of factors that could explain this heterogeneity include not only gender, age, and career experiences but also socialization, reasoning styles, and emotional state—all traits that could affect the outcome of strategic interactions.

Although this research program is still in its early stages, the focus on individual heterogeneity offers several big payoffs. The first is more empirically realistic models of individual decision-making processes. As Powell argues in his conclusion to this issue, this might be done either by relaxing rationalist assumptions about how states act or by devoting more attention to actual decision makers. The resurgence of the study of leaders in international relations exemplifies this latter strategy and is visible in a number of contributions to the special issue.[22] The focus on individual-

19. Bayram 2011, 2016a, 2016b; Herrmann and Fischerkeller 1995.
20. Kertzer and McGraw 2012.
21. Hafner-Burton et al. 2014.
22. For examples beyond the special issue, see Byman and Pollack 2001; Colgan 2013; Croco 2011; Debs and Goemans 2010; Hall and Yarhi-Milo 2012; Horowitz, McDermott, and Stam 2005; Saunders 2009; Weeks 2012.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

level heterogeneity also opens up opportunities to connect with a wider body of behavioral research in political science, including studies of public opinion. A number of contributions to this issue pursue this approach.

The articles we've assembled explore at least four such sources of heterogeneity. The first is differences in resolve and attitudes toward power, emerging in part out of a literature on time preferences.[23] A second body of work considers social preferences, including the nature and depth of in-group versus out-group identifications.[24] This research exhibits affinities with constructivism and can open a bridge between empirical behavioral research and constructivist theories of preference formation. Third, we consider work on how mass publics absorb and use information, contrasting a variety of biases with standard economic motivations.[25] And fourth, we look at how differences in experience affect the preferences, beliefs, and decision making of elites and particularly their capacity to think strategically.[26] These sources of heterogeneity by no means exhaust existing and potential work at the intersection of international relations and the new behavioral revolution but they offer a sense of the scope of the advances that are within grasp.

A second major theme opened, but certainly not solved, by the behavioral revolution is what we call the aggregation issue in international relations: how we move from individual to collective decision making. One way to finesse the aggregation issue—and one defended by game theorists—is to impute preferences to the nation state and derive predicted outcomes accordingly; we consider how this approach might actually be used in the context of the findings of the behavioral revolution.[27] The alternative is to incorporate psychological considerations more directly into the analysis of decision-making processes. Early work on these questions, especially studies of "group think," considered how decision-making processes can *amplify* individual-level bias.[28] However, as we argue in more detail and Saunders shows in her contribution, institutions are designed in part precisely to *mitigate* such biases—for example, by lowering the costs of obtaining information and ensuring that competing views are represented. The behavioral revolution in international relations raises long-standing questions about decision making, but in the context of new models of individual motivation and behavior.

To develop our arguments about the gains from bringing the behavioral revolution into international relations,we first survey core features of the rationalist approach to international relations, with particular attention to underlying assumptions related to preferences, beliefs, and decision making. Second, we outline the core elements of the behavioral revolution itself, again focusing on clusters of research related to

23. See articles by Kertzer; Rathbun, Kertzer and Paradis; Renshon, Lee, and Tingley; Tingley in this special issue.
24. See Bayram and Herrmann, both in this issue, and Herrmann 2009.
25. See Rho and Tomz 2017.
26. See Saunders 2017.
27. See Powell 2017.
28. Janis 1972.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

nonstandard preferences, beliefs, and decision making. Next we explore the insights to be gained by focusing on empirically derived dimensions of individual-level heterogeneity, including the extent to which decision making does in fact conform with rationality. We suggest a research agenda for the analysis of institutions and decision-making processes that incorporates the insights of the behavioral revolution. Finally, we then raise two ongoing methodological critiques of the experimental research that has driven the behavioral revolution: external validity and replication.

The Rationalist Approach to International Relations

There is no single, canonical model of cooperation and conflict in international relations. Nonetheless, all rationalist theories of international relations rest on important assumptions about the environment and the actors.[29] Individuals are assumed to maximize expected utility by determining the payoffs attached to all possible outcomes, assessing their probabilities, updating information on those probabilities, and choosing the strategy with the highest expected return. In game-theoretic models, equilibrium outcomes are generated out of the choices of two or more players in the game. Rationalist models, in turn, can be decomposed into environments and actors, with two sets of variables in each category.

Strategic environments are composed, first, of the actions available to the actors, formally represented by a game tree of the different paths along which events might unfold. The environment is also composed of an information structure that defines what the actors know and what they have to infer, often imperfectly or at high cost, from the behavior of others. The environment is treated in a very similar way in the new behavioral work. Actors still choose between alternative courses of action and analysts must specify what those actors know about others' actions or preferences.

Within the rationalist approach, actors are viewed through the lens of their preferences and beliefs. Differences in preferences and beliefs provide ample room for a diversity of modeling choices. Preferences are simply the actor's subjective rank-ordering of the terminal nodes or outcomes of the strategic interaction, and are assumed to be transitive. Preference rankings are always with respect to some conception of value or utility held by or imputed to the actors in question.[30] In the field of international political economy, for example, the preferences of interest groups and households are assumed to center on welfare, often defined in terms of real income; for firms it is profitability. By contrast, the field of security studies has tended to focus on stylized preferences such as the distinction between status-quo and revisionist powers or between "doves" and "hawks" who differ in their tolerance for conflict. Within this tradition, the role of individuals has played a role.

29. This approach is discussed in more detail in Lake and Powell 1999.
30. Frieden 1999; Moravcsik 1997.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Many scholars have modeled the preferences of agents who make foreign policy decisions, including bureaucracies, leaders, or even the proverbial median voter.[31] The theory of "audience costs" provides a well-known example of this variety of theorizing.[32]

The second core feature of actors is their beliefs or what actors assume about the state of the world, including the preferences of other actors. Models of complete information make very demanding assumptions about the correspondence between what players believe and the actual state of the world. For example, players are assumed to have common knowledge about their environment, as well as the preferences of other actors and the actions they have taken.[33] Games of incomplete information, by contrast, assume only that players have common knowledge over the structure of the game and beliefs about how parameter values are distributed.[34] Such models have played a particularly important role in international relations theory, for example, in signaling games in which parties are uncertain over the other's true type but can take costly actions to reveal information about their own.[35]

Tying the strategic environment and the attributes of actors together are decision-making processes that can operate at the level of individuals or various collectives such as electorates, committees, parliaments, and informal "inner circles" of top decision makers. Decision making entails evaluating the environment and ordering possible outcomes in light of preferences and beliefs about the state of the world. By assuming that actors maximize expected utility and are rational, the approach posits that, on average, decisions are "optimal." Actors correctly understand their environment and update their beliefs appropriately, given the information available to them. And given the environment, their preferences, and their beliefs, actors make choices that will, probabilistically, return the greatest utility.

What's New in the Behavioral Revolution: Preferences, Beliefs, Decision Making

The behavioral revolution grew out of a suspicion that expected utility theory had fundamental limitations as a positive—as opposed to normative—theory of human behavior. The revolution had important precursors in the work of Herbert Simon

31. For example, Allison 1971; and Horowitz, McDermott, and Stam 2005, respectively.
32. Fearon 1994; Tomz 2007; Weeks 2008.
33. This common-knowledge assumption not only requires that players know these variables, but also that they know other players know them, and know that other players know that other players know them as well … ad infinitum. Aumann 1976; Geanakoplos 1992; Rubinstein 1989.
34. For example, even if a player is uncertain whether an opponent possesses preferences characteristic of a prisoners' dilemma or "Chicken" game, she nonetheless holds a prior belief about how those preferences are distributed and acts accordingly. Aumann and Brandenburger 1995; Harsanyi 1967.
35. Banks 1990; Fey and Ramsay 2011; Wagner 2000.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

on bounded rationality and the identification of a number of anomalies in expected utility theory.[36] But the origins of the revolution are usually traced to the joint work of Daniel Kahneman and Amos Tversky, and particularly to two articles. One used psychological models of the brain as an information-processing device to argue that humans are not good at statistical reasoning. Rather than maximizing expected utility, Kahneman and Tversky found that individuals used heuristics that led to substantial and pervasive over- and underestimation of probabilities.[37] The second contribution, their seminal work, outlined the three pillars of prospect theory: a) its alternative decision-making model in which heuristics and reference points played a key role; b) the treatment of decision making as an empirical question that should be grounded in experimental evidence; and most strikingly c) the argument that actual human decision making frequently violates axiomatic elements of expected utility theory. In particular, they found that decision makers assessed utility in terms of gains and losses from a reference point, rather than in terms of absolute gains, and that the value assigned to losses and gains of equal magnitude was in fact asymmetric.[38]

As Stein outlines in her concluding essay, prospect theory generated a host of new models and experimental evidence from the mid-1980s. Much of this work involved single-agent choice experiments that studied anomalies such as the effects of framing, reference points and endowment effects, intransitive behavior, systematic under- or overweighting of probabilities, hyperbolic discounting, and violations of Bayes theorem in the form of failure to update expectations in the face of contradictory evidence.[39] This research also used strategic interaction experiments. Of particular interest to game theorists were studies of social preferences (e.g., preferences for fairness, aversion to inequality, and altruism) that affected strategic outcomes, including the propensity to cooperate.[40] The result of these innovations was the birth of behavioral game theory.

The behavioral revolution did not begin with a strong theory of cognition, but one gradually emerged out of this work and has ultimately been given a biological foundation in new brain research.[41] This approach conceived of cognition as consisting of two systems: an intuitive "System 1" that is quick and cognitively efficient and therefore typically accounts for most decision making. However, it is vulnerable to the

36. Allais 1953; Strotz 1955. The first wave of psychological research in international relations, exemplified by Jervis 1976, drew largely on this earlier strain of theorizing. We discuss this earlier strain, largely under the heading of beliefs.

37. Tversky and Kahneman 1974.

38. Kahneman and Tversky 1979. Another important early paper was Richard Thaler's (1980), which noted similar anomalies in the value that respondents would attach to goods that they owned and didn't own, a finding he called the "endowment effect." See also Tversky and Kahneman 1992.

39. El-Gamal and Grether 1995; Frederick, Lowenstein, and O'Donoghue 2002; Kahneman, Knetsch, and Thaler 1990; Kahneman, Slovic, and Tversky 1982; Laibson 1997; Loewenstein and Elster 1992; Loomes, Starmer, and Sugden 1991; O'Donoghue and Rabin 1999; Thaler 1999.

40. See Akerlof and Yellen 1990; Fehr and Schmidt 1999; Rabin 1993.

41. Particularly Bechara, Damasio, and Damasio 2000.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

catalogue of biases we outlined. A new round of work in the 1990s augmented this model of System 1 by considering the role of emotion in decision making.[42] The second, more slow-moving System 2 is deliberative and more closely approximates rational choice. Because System 2 demands effort and is therefore difficult, reliance on it is distributed unevenly across any population.

At its inception, many advocates for a behavioural approach to studying decisions had the highly subversive objective of identifying universal deviations from rational choice. A drawback of this approach was a laundry list of biases, but with the corresponding difficulty of knowing when such deviations would arise, with what magnitude, and whether they were in fact constant across agents. Since those early experiments, however, progress has occurred by focusing on how systematic variation in preferences, beliefs, and decision making affects the outcome of standard choice and strategic problems; we focus on each of these three pillars of rational choice models here.

Preferences

Rationalist models allow for a diversity of preferences but the behavioral revolution has identified three potential deviations from rationalist assumptions that are especially relevant for international relations: how actors actually assess risk; how they discount the future; and the possibility of social preferences.

Although rationalist models allow for heterogeneous risk preferences, prospect theory suggested that such preferences might be structured in predictable ways. Individuals tend to make choices not with respect to total returns but with respect to deviations anchored in the status quo. Experimental research also showed that subjects were risk averse with respect to gains but risk acceptant with respect to losses. For international relations scholars, the applications of these findings to international security were immediately evident.[43] Stein's account of Sadat's behaviour in 1973 explained the decision for war not as an irrational decision but one anchored in Egypt's prior possession of the Sinai.[44] In the domain of losses, Sadat was much more risk acceptant than he might have otherwise been. Israeli decision makers, in turn, adopted the wrong model to predict Sadat's behavior and were caught badly off guard. Such findings have clear implications for the stability of deterrence, which will depend not only on the balance of capabilities but also on how actors perceive the status quo.[45]

Second, the behavioral revolution has shed insight into how actors discount the future.[46] Standard models allow for a range of discount rates but assume that

42. See Haidt 2012, for example, on the role of disgust.
43. Davis Jr. 2000; Farnham 1994; Jervis 1992; Levy 1992a, 1992b, 1997; McDermott 1998; Stein and Pauly 1992.
44. Stein 1985.
45. Berejikian 2002; Mercer 2005.
46. Coller and Williams 1999; Fowler and Kam 2006.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

actors apply the same discount factor when comparing costs and benefits that arise between any two time periods. By contrast, experimental research suggests pervasive time inconsistency or "hyperbolic discounting" problems.

The applications to international relations are wide ranging, but have yet to be explored at length. The most obvious example is negotiations over climate change. Countries have agreed on bold goals, such as stopping warming at 1.5 or 2 degrees Celsius above pre-industrial levels, yet they pursue policies likely to cause perhaps double that warming by 2100 and even more in the years beyond. Of course, these outcomes might be modeled in a standard collective action framework, through the lens of electoral constraints on requisite policy reform or as a consequence of intergenerational conflicts. But they may well be rooted in underlying citizen and elite preferences characterized by hyperbolic discounting.

Hyperbolic discounting has a wide array of other possible effects on existing models in international relations. For example, such behavior renders strategies of reciprocity within the Prisoners' Dilemma less likely and less effective.[47] However, it may make bargaining problems easier to resolve precisely because the shadow of the future weighs less heavily on the present.[48] For example, variation in the extent to which decision makers discount the future helps explain preferences for negotiating and joining international agreements.[49]

Third is the question of individual (pro-self) versus social (altruistic or sociotropic) preferences, an issue that provides wide opportunities to engage the constructivist research agenda. Although rationalist models allow for the possibility of social preferences, standard game-theoretic models typically assume that actors are strictly egoistic and thus are prone to free ride in games of collaboration. This may seem reasonable in the anarchic, self-help environment of international relations. However, a particularly well-developed body of both theory and experimental research showed quite early that people who played public goods games were more cooperative than the standard theory would predict; participants were even shown to punish free riders.[50] When provided with an opportunity to be altruistic or self-regarding, people are more inclined to share and trust than canonical rational theory would predict.[51] These tendencies may help explain why, at least in some domains of international behavior, self-interested states join international institutions in large numbers and comply with expectations even in the absence of formal enforcement mechanisms.[52]

47. Contrary to Axelrod 1984.
48. Fearon 1998.
49. Edelstein 2002; Hafner-Burton et al. 2014; Krebs and Rapport 2012.
50. For example, Ledyard 1995; Rabin 1993; for an early review, Andreoni 1995; Fehr and Gächter 2000; Fischbacher, Gächter, and Fehr 2001. Similar results are evident in other kinds of cooperation and sharing games such as the ultimatum, dictator, and trust games, and in a variety of empirical settings. For a review of findings in political science, see Wilson 2011; for an example in international relations, LeVeck et al. 2014.
51. For work on trust in international relations more generally, see Kydd 2010 and Rathbun 2009, 2011.
52. Hafner-Burton, LeVeck, and Victor 2016b.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Beliefs

In full information models, actors not only have correct views of the state of the world but also about the preferences and beliefs of other players in the game as well. Imperfect and incomplete information models relax this assumption but nonetheless assume that beliefs are formed and updated in a way that avoids systematic error.

The behavioral revolution has challenged these assumptions about beliefs—especially where updating beliefs about the world requires obtaining and interpreting large amounts of information and making complex inferences.

First and most fundamentally, assumptions about beliefs must deal with the pervasive problem of misperception. Since the 1950s, psychologists have known that misperception is not a random event; certain kinds of risks and opportunities invite actors to embrace erroneous beliefs about the world. These ideas found traction in international relations through the influential work of Robert Jervis who focused on when, how, and why actors misunderstood the actions of their opponents, such as incorrectly perceiving intended cooperation as hostility.[53] An important strand of subsequent work focused on how beliefs about the world were shaped by available heuristics, metaphors, and analogies, even when inappropriate to the strategic setting.[54]

The new behavioral revolution offers fuller insight into beliefs, starting with the failure to draw appropriate inferences, including statistical inferences. These problems include overweighting of available evidence, ignoring information that is harder to obtain ("what you see is all there is"), and failure to pay attention to the base rate of any given phenomenon and thus wrongly estimating the probability of its future occurrence.[55]

The most wide-ranging investigation of weak statistical inference in foreign policy decision making is Philip Tetlock's analysis of foreign policy experts' forecasting capabilities.[56] The core finding of Tetlock's work is that experts, especially those who have strong ideologies, are surprisingly poor forecasters. They overweight single modes of causation as well as favored outcomes rather than pay attention to the frequency of confounding events. As a result, they often tend to overestimate threats because they can imagine the causal pathways to conflict. A prominent example is the widely cited statement by former Vice President Richard Cheney that "if there's a 1 percent chance that Pakistani scientists are helping al-Qaeda build or develop a nuclear weapon, we have to treat it as a certainty."[57] Such a response is equivalent to overweighting the probability of a nuclear terrorist attack and thus overinvesting in responses to this threat while, in a world of scarce resources, necessarily underinvesting in other more probable challenges.

53. Jervis 1970, 1976.
54. Hemmer 2000; Houghton 1998; Khong 1992; Neustadt and May 1988; Reiter 1996; Shafer 1988.
55. For example, Barberis, Shleifer, and Vishny 1998.
56. Tetlock 2005; Tetlock and Gardner 2015.
57. Quoted in Suskind 2007, 62.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Recent work has extended this analysis into how actors model strategic environments.[58] Experimental studies show that people do not generally iterate to equilibrium, perhaps because they think that opponents do not iterate games fully either.[59] But recent studies in neuroeconomics also show that more strategic reasoners are more cooperative in repeated assurance games, precisely because they anticipate their adversaries' likelihood of cooperating in response to their own cooperation.[60] Given the central role of beliefs for strategic interaction in international relations, these insights have broad implications for both theory and empirical work, yet only a very small number of studies have used experimental techniques to test these insights.

Finally, it is important to emphasize that misperception is not limited to the strategic setting or adversaries but extends to misperception of one's *own* capabilities and the corresponding problem of overconfidence. It has long been recognized in international relations that overconfidence is a cause of bargaining failures and war.[61] Experimental research has subsequently found that the problem is ubiquitous. For example, experiments have shown that the vast majority of drivers think their skills are above average.[62] Elite business school students think they are much better at forecasting economic activity than they actually demonstrate in real-world contests.[63] As we implied earlier, such self-delusory beliefs may have been responsible for key decisions in the Iraq War.

Decision-Making Processes

Even given a rationally ordered and well-behaved set of preferences and correct beliefs about the world and other players, individuals' and groups' decision-making processes may skew decisions away from the predictions of rationalist models. This idea has a long tradition in international relations, influenced at the outset by models of bounded rationality from Herbert Simon and others.[64] Constraints on information processing and cognitive shortcuts and heuristics at the individual level can systematically bias decision making away from the predicted equilibria of expected-utility models.[65]

What is new in the more recent studies of the behavioral revolution is not only an improved understanding of cognitive shortcuts but a deeper understanding of

58. Camerer, Ho, and Chong 2004; Costa-Gomes and Crawford 2006; Stahl and Wilson 1995. Experimentally, the diversity of ways in which actors may model the strategic setting has been studied extensively using "beauty contest" models. Nagel 1995.
59. Camerer, Ho, and Chong 2004; Costa-Gomes and Crawford 2006; Stahl and Wilson 1995.
60. Yoshida et al. 2010.
61. Blainey 1973.
62. Svenson 1981.
63. Alpert and Raiffa 1982.
64. See Conlisk 1996; Rubinstein 1998; and Simon 1955.
65. Janis 1972; Jervis 1970, 1976; Steinbruner 1977.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

decision making more generally, including the relationship between reason, intuition, and emotion. It may be plausible to assume that decisions reflect optimization of expected utility when the stakes are large and leaders have ample organizational resources to bring to bear on the problem at hand. Yet even in those situations there can be large influences on decisions from deficits in attention—especially for busy elites—as well as inappropriate selection of heuristics and emotions.

The list of decision-making biases is large, and one frontier in research is to understand how these biases affect the kinds of strategic interactions that rationalist models capture. One cluster of behavioral research looks at how decision makers confront complex menus of choices, a common attribute of both routine and crisis foreign-policy decision making. In such settings, experimental evidence suggests that the framing of choices has a strong influence over decision making: decision makers prize the familiar, available, or salient.[66] This simple point explains intense battles for agenda control when complex political choices are at stake. Decision makers also avoid choices that require complex deviations from the status quo, even when warranted by the circumstance; the slow path to a counter-insurgency strategy in Vietnam and Iraq provide examples.[67] Such biases might also account for the familiar but undertheorized "lock-in" effect of international institutions.[68]

We are mindful that the channels through which psychological and behavioral factors operate are complex, and that the boundaries between how they affect preferences, beliefs, and decision making are not sharp. Dispositions can drive change in all three of these attributes in specific individuals. The impact of emotion is a striking example.[69] Herrmann shows that emotions affect preferences, for example, by linking individual interests to the collective. But emotion also affects beliefs about allies and adversaries and decision-making processes as well.[70] Despite these fuzzy boundaries, there is an advantage in focusing on the distinct effects of the behavioral research program on our understanding of preferences, beliefs, and decision making because it permits dialogue with existing theory.

The Sources and Consequences of Behavioral Heterogeneity

As we noted, behavioral research in international relations has shown an interest in two major themes that differentiate it somewhat from other disciplines. The first is a focus on the effects of actor heterogeneity among both elites and masses. In the next section we focus on a second major theme: the influence of psychological and behavioral factors on collective decision-making processes and the process of

66. DellaVigna 2009.
67. Lake 2010–11; Shafer 1988.
68. Ikenberry 2001; Krasner 1976, 1983.
69. Bleiker and Hutchison 2008; Hall and Ross 2015; Hymans 2006; Lieberman, Schreiber, and Ochsner 2003; Mercer 2010; Neuman et al. 2007.
70. See also Ho and Imai 2008; Huberman 2001.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

TABLE 1. *Examples of Nonstandard Preferences, Beliefs, and Decision-Making Procedures*

Category	*Trait*	*Exemplary work from behavioral sciences*	*Existing applications in IR and possible extensions*
Nonstandard Preferences			
	Risk and Prospect Theory	Kahneman and Tversky 1979	Irrational territorial expansion (e.g., Stein 1985); cooperation to cement gains in the status quo; deviations from rational choice
	Time discounting	Coller and Williams 1999; Laibson 1997; Zimbardo and Boyd 1999	Differences in time-weighted behavior between short-horizon democrats and long-horizon autocrats
	Self-regarding versus social-regarding preferences (e.g., cooperativeness)	Fowler and Kam 2007; Haigh and List 2005; Lopez, McDermott and Petersen 2011; Plott and Zeiler 2005	High levels of cooperation in international affairs; compliance with international agreements; declining levels of violence
	Ambiguity aversion	Halevy 2007; Trautmann and Zeckhauser 2010	Bias toward familiar policy choices, even if suboptimal
	Inequality aversion	Engelmann and Strobel 2004; Fehr and Schmidt 1999	Foreign aid
Nonstandard Beliefs			
	Misperception: application of faulty heuristics to explain strategic situations	Camerer, Ho, and Chong 2003; Chi, Glaser, and Rees 1982; Simon 1955	Reasoning by wrong analogy—for example, viewing Iraq through the Vietnam lens
	Misperception: strategic opponents	Camerer, Ho, and Chong 2002; Nagel 1995; Stahl and Wilson 1994	Failure to make strategically rational moves, leading to more cooperation in the face of defection risks (e.g., Hafner-Burton et al. 2014)
	Overconfidence	Alpert and Raiffa 1982; Johnson and Fowler 2011 Massey and Thaler 2010; Svenson 1981	Poor forecasting, especially by experts (e.g., Tetlock 2005)
Nonstandard Decision-Making Procedures			
	Heuristic rather than analytical reasoning	see "Misperception"	Failure to anticipate "obvious" errors in decision making, such as failure during period of rapid-fire, attention-deprived decisions to consider security needs in postwar Iraq (Lake 2010–11)
	Poor choices from complex menus	Ho and Imai 2008; Huberman 2001	Avoidance of choices that require complex deviations from the status quo—for example, the slow realization of the importance of counter insurgency in Iraq and Vietnam (Shafer 1988); lock-in effects of international institutions (Krasner 1976; Ikenberry 2001).
	Misperception: information cascades and other groupthink	Alevy, Haigh, and List 2007	Tendency for groups to reinforce poor decisions (e.g., Allison 1971) and inability of complex organizations to engage in strategic behavior
	Emotion ("hot versus cold" reasoning)	Ho and Imai 2008; Lieberman, Schreiber, and Ochsner 2003; Schwarz 2000	Difficulties in signaling and developing reputation (e.g., Mercer 2005); escalation during conflict (e.g., Gries 2004); extreme responses during war on terror (e.g., Crawford 2009); ideological responses to uncertainty (Kertzer and McGraw 2012)

Sources: Excerpted from DellaVigna 2009; Hafner-Burton, Hughes, and Victor 2013; and main text.

aggregation. In both sections, the articles in this special issue provide examples of the new research program emerging in international relations.

Formal rationalist models are typically built around actors of different "types," exhibiting different preferences or beliefs about the world. The behavioral revolution can ground this delineation of types more closely in empirical research. Among the principal dimensions of variance emerging in the literature, and central to the articles in this special issue, are the extent to which actors have egoistic as opposed to social preferences (Bayram; Herrmann; Rho and Tomz), differences in attitudes toward power and resolve (Kertzer; Renshon, Lee, and Tingley; Tingley) and variation in the nature and role of experience and in the capacity to think strategically (Saunders; Rathbun, Kertzer, and Paradis).

Herrmann exemplifies new work on social preferences. He claims that national identities shape beliefs that in turn can license choices driven by emotion rather than strategic calculation. Individuals with more intense national attachments are more likely to attribute malign intentions to countries that are disliked and benign intentions to those that are liked, even when these states are portrayed as engaging in exactly the same behavior. This attribution bias, in turn, leads to distorted beliefs about the frequency and nature of international conflict and influences individual attitudes toward particular conflicts, including the Iraq war. Once he was characterized as hostile, Saddam Hussein could do little that was likely to change beliefs for many Americans, including key decision makers.

In this volume, Rho and Tomz advance recent work on the effects of altruism and show they are contingent on knowledge and information.[71] They focus on how altruism leads to variations in the desire to help society in general and the poor in particular. Like others in the literature, they find that individuals support trade protection that is expected to help others even if it hurts their own economic welfare. In a novel experiment on the effects of economic knowledge, however, they provide respondents with information on the winners and losers from trade barriers. They find that most individuals do not understand the economic consequences of protectionism. However, when informed about the general welfare-enhancing effects of trade openness, more educated respondents become more self-interested while less-educated respondents become more altruistic. They conclude that even as the level of altruism may vary across the population, its effects on policy preferences are contingent on individual knowledge and information.

Alongside the study of trade preferences is an emerging literature applying insights from the behavioral revolution to debates over compliance with international commitments more broadly.[72] Bayram, in this issue, draws on the literature on social identity and social preferences to examine the puzzle of why some politicians are motivated by a sense of legal and normative obligation toward international law while others

71. Other work on social preferences in international political economy includes Guisinger 2014; Prather 2014; and Sabet 2014. Out-group feelings were first explored in Mansfield and Mutz 2009.

72. Galbraith 2013; Hafner-Burton, LeVeck, and Victor 2016a, 2016b; Poulsen 2013.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

aren't. She argues that legal obligation is rooted in cosmopolitan social identity. Linking social identity theory to the "two-level model" of cognition, she hypothesizes that the degree of cosmopolitan identification explains the heterogeneity in decision makers' feelings of legal obligation, which in turn affects how actors respond to choices about compliance. In an original survey of German parliamentarians that is replicated in a sample of college students, Bayram finds that cosmopolitans are "dutiful compliers." Noncosmopolitans, by contrast, employ a cost–benefit approach to compliance; they are "instrumental compliers." The cosmopolitan group is less sensitive than the noncosmopolitans to the material payoff structure of the compliance game.

The analysis of social preferences provides an obvious foundation for studies of cooperation. But the behavioral revolution also has important things to say about power and conflict as well, particularly through a consideration of time preferences. Drawing an analogy between the international relations concept of resolve and the psychological literature on willpower and time preferences, Kertzer (in this issue) considers the long-standing puzzle of why some states display remarkable persistence in war while others "cut and run." Using laboratory experiments on the onset and duration of a hypothetical military intervention, he considers the effects of participants' time and risk preferences, measured using techniques borrowed from behavioral economics. More patient individuals are less sensitive to casualties—perhaps because they place greater value on the long-term effects of military intervention. More risk-averse individuals are more sensitive to the human costs of "staying the course," but also to the reputational costs of withdrawal. Kertzer shows how individuals with different time preferences construct very different mental representations of the costs of war even when facing situations characterized by identical structures and incentives.

In addition to studies of resolve, new literature is also reassessing how individuals think about changing power, a crucial variable in models of conflict that rest on the credibility of commitments. Using a series of experimental vignettes on declining US and rising Chinese power, Tingley's contribution to this issue finds high variation in how actors reason about power transitions. He demonstrates that individuals respond to changes in power in ways consistent with prospect theory, where decisions are heavily anchored by evaluation against the status quo. Moreover, individuals view a change in their own country's power very differently from a change in a possible competitor, a finding consistent with Herrmann's contribution. Tingley also makes an important methodological contribution, showing that open-form, unstructured questions elicit very different information about individual preferences, beliefs, and decision making than standard closed-form surveys that reflect researchers' expectations.

Renshon, Lee, and Tingley contribute to a growing literature on the effect of emotion, pushing the analysis of responses to changing power down to the biochemical level by using physiological measures of emotional arousal. They demonstrate that such arousal leads individuals to make strategically suboptimal decisions, at least as formally modeled in games with credible commitment problems. They

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

suggest that "aversive physiological arousal" short-circuits subjects' ability to think strategically and in systematic ways. These results conform closely to the two-system models of decision making, suggesting that stressed individuals are more likely to rely on heuristics and make quick System 1 decisions.

A final dimension of heterogeneity that is garnering attention is the very capacity to think strategically. In this issue, Rathbun, Kertzer, and Paradis argue that rationality is a variable rather than a constant, rooted in two psychological characteristics. Individuals with strong "pro-self" or egoistic value orientations and strong "epistemic motivation," or the desire to think strategically, are more likely to act "as if" rational. However, this type is by no means ubiquitous; to the contrary, many people are lacking in one or the other trait. Using both laboratory experiments and a historical case study, they examine how games played between as-if rational and less-than-rational agents generate equilibria that do not conform to standard conflict models. In a novel interpretation of German foreign policy in the 1920s, they suggest that the inability of German leaders to reason strategically produced overreach and a breakdown in diplomacy that set the stage for future conflict.

Highly experienced elites are more likely to exhibit the attributes of rational decision making, including skilled strategic bargaining, even if they are also more prone to overconfidence.[73] Individuals who can reason "down the game tree" at lower cost or with less cognitive effort advocate different foreign policy strategies. For example, Hafner-Burton and colleagues use experiments drawn from behavioral economics and cognitive psychology to show that patient decision makers are more likely to prefer complex treaties with larger numbers of countries (and larger long-term benefits).[74] Most significantly for a number of foreign policy issues, new research is showing not only differences among elites but particularly sharp differences in how foreign policy elites and average citizens play games.[75]

In this issue, Saunders explores how a leader's level of foreign policy experience along with small-group decision-making structure affect decision making. The study examines US decision making in the Persian Gulf War of 1991—with an experienced leader in George H.W. Bush—and the Iraq War of 2003, under the leadership of the less-experienced George W. Bush. Drawing upon the effects of expertise on individual decision making, and extending insights from principal-agent theory, Saunders hypothesizes that the balance of experience between a leader and his advisers can magnify or diminish cognitive biases. Inexperienced leaders are more dependent upon yet less effective at monitoring the behavior of experienced advisers. They are less able to diversify the advice they receive, and more likely to prefer policies that appear certain. All of these tendencies of inexperienced leaders push policy in a "riskier" direction, making war more likely by failing to create incentives to gather and analyze information adequately.

73. Hafner-Burton, Hughes, and Victor 2013.
74. Hafner-Burton et al. 2014.
75. Fisman et al. 2015; LeVeck et al. 2014.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

It is important to underline that standard models are also built on the assumption that actors have diverse preferences; these are explicitly modeled as such in the divergent preference orderings in game-theoretic models. What is new in behavioral models is a substantial widening of the dimensions of heterogeneity, an empirical grounding of those differences, and a deeper appreciation of the extent to which they have psychological roots. In some cases, such as with respect to differences in resolve, these new insights resonate with existing literature, such as the age-old distinction between hawks and doves. In other cases, however, as with discussions of motivated reasoning, time, or social preferences, the differences across actors may require rethinking the axiomatic foundations of rational choice itself.

Aggregation: The Behavioral Revolution and the Study of Collective Decision Making

Nearly all the models of nonstandard preferences, beliefs, and decision making focus on the individual level. They are thus immediately useful for understanding policy preferences, the political behavior of citizens (such as voting with respect to foreign policy issues), and elite decision making. However, as Powell argues forcefully in his conclusion, most rationalist models in international relations are models of *state* behavior. How do we get from the new behavioral findings at the individual level to collective decision making?

Powell argues that behavioral approaches have two analytic choices. The first is to simply attribute nonstandard preferences, beliefs, and decision making directly to states, as rationalist approaches do. The alternative, favored by all of the authors in this issue, is to maintain the commitment to analysis at the individual level, while addressing how individual-level findings can be aggregated to understand collective as well as individual decision making.

Of course, not all rationalist approaches are committed to the "states as actors" framework that Powell favors. A growing strand of rationalist theory adopts what might be called a "boxes-within-boxes" framework to address the aggregation process,[76] and the articles here have generally taken the same approach. In such models, political processes in one "box" or level are modeled but then treated as inputs into political processes in a second "box." This strategy allows analysts to focus on the parts of the complicated causal chains running from individual preferences to strategic interactions at the international level that are deemed causally significant for the outcome in question. For example, the democratic peace literature posits that macro-level institutional arrangements—particularly regime type—are central to questions of war and peace; interest groups and coalitions are seen as secondary. Conversely, many open-economy politics models focus on underlying citizen

76. Lake 2009; Lake and Powell 1999.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

preferences and resultant interest groups as the drivers of policy, with institutions acting as "cash registers" that simply sum the vector of these interests.

The contributors to this special issue have focused on different steps in this boxes-in-boxes approach and different approaches to how research on individuals can be aggregated. Kertzer, Rho and Tomz, and Herrmann focus almost exclusively on the preferences of broad publics with respect to resolve during crises, trade policy, and broader foreign policy choices respectively. They do not explicitly address how preferences are aggregated, but they rightly note that these preferences are intrinsically interesting because they act as constraints on national decision makers in democratic settings. They suggest a variety of ways in which this might work, including constraints associated with public opinion and the electoral process.

A second cluster of articles, including Tingley; Renshon, Lee, and Tingley; and to some extent Herrmann as well, treat publics as samples of convenience for understanding how leaders make decisions. They implicitly argue that we cannot simply impute rational action to leaders—however constrained by the strategic setting—but must justify that individuals are likely to respond in ways that are consistent with postulated models. They do so on the grounds that the theories of behavior in question are likely to be general in nature. For example, Tingley examines whether individuals appraise shifts in power in ways that are consistent with the credible commitment logic. He finds that there is in fact tremendous heterogeneity in how individuals understand such settings and infers that elites may not necessarily behave as such models expect either, for example, showing a commitment to consistent positions despite changes in power or focusing not on the effects of changing power over time but on cost–benefit calculations.

A third cluster of articles—including Bayram; Rathbun, Kertzer, and Paradis; and Saunders focus directly on elites and the heterogeneity across them. As with other domestic models of foreign policy, they expect the behavior of similarly situated states to vary but not only on the basis of factors such as regime type, partisanship, domestic coalitions, or differences in audience costs. Foreign policy choices will also differ depending on factors such as the social preferences of decision makers or their cognitive motivation.

The next frontier for this sort of work is to think more deeply about how to integrate the new behavioral findings into the study of institutions—both large and small—that actually generate foreign policy and bargaining choices. Saunders's contribution suggests three promising fronts for further theoretical and empirical work: the long-standing program on the effects of individual-level characteristics and heterogeneity on decision making; the effects of institutional design on bias and error; and the role of social preferences in decision making.

First, individual heterogeneity may help explain why the same institutions can yield such disparate outcomes, quite apart from more standard rationalist explanations. Leaders and other political actors do not simply differ on factors such as partisanship or imputed preferences, such as being hawks or doves; they may also be more or less experienced, more or less able to think strategically, or more vulnerable to emotional appeals. This theme of elites' individual-level heterogeneity figures

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

prominently in a number of the articles in the project and reflects the revival in the systematic study of leaders in international relations more generally.[77]

Second, institutions are typically designed to bias or constrain policy in particular ways.[78] It is precisely the role of institutions to privilege certain actors and their interests in the policy process; that is why fights over them can be so heated.[79] New work on institutional design for arbitration in investment law and dispute resolution in trade law, for instance, shows that these institutions bias the role of individuals—arbitrators and dispute panelists—in differing ways.[80] Cognitive biases introduce another layer of considerations to our understanding of institutional choice. For example, it is hard to explain the status quo bias or "stickiness" of institutions in fully rationalist terms.[81] If altering institutions is a complex political process with uncertain outcomes, and cognitive effort is costly, psychological factors may help explain observed status quo bias. Once formed, groups with common interests are resistant to change not because of their material interests but because change is difficult and costly even if new rules would be more efficient.

However institutions may also be designed and used precisely to correct individual bias or offset cognitive limitations, as Bendor and Hammond show in their influential critique of Allison.[82] Bureaucratic procedures are often designed to collect and disseminate information, evaluate options, and study the consequences of choice more systematically and thoroughly than a single and especially "intuitive" decision maker can do. Contestation between rival departments can also lead to the airing and resolution of differing views, leading to more "rational" decisions. What appear to be slow, dysfunctional, and contentious procedures from a rationalist perspective may be quite effective and appropriate if we begin with the assumption that individuals working within complex organizations are vulnerable to cognitive biases, misperception, mistaken beliefs, and outright error. Bureaucratic procedures in effect transform possible System 1 decisions at the level of the individual into slow, agonizing, but more rational System 2 policies at the level of the aggregate. This suggests the counterintuitive possibility that policy can be more "rational" than any of the individuals involved, at least in settings where institutions successfully incentivize robust decision-making processes.

Finally, the behavioral revolution may even speak to the long-standing question about whether it is possible to speak about a "national interest" without reification. In rationalist terms, it is difficult to imagine any real aggregation process that would produce an interest different from, or greater than, those of contending groups and parties. However, as suggested in Bayram's contribution to this issue,

77. Byman and Pollack 2001; Colgan 2013; Croco 2011; Debs and Goemans 2010; Hall and Yarhi-Milo 2012; Horowitz, McDermott, and Stam 2005; Saunders 2009; Weeks 2012.
78. Rogowski 1999.
79. Gourevitch 1999.
80. Pauwelyn 2015.
81. Jupille, Mattli, and Snidal 2013.
82. Allison 1971; Bendor and Hammond 1992.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

the kinds of sociotropic preferences identified in the cognitive revolution open the door to just such a generalized understanding of collective interests, and one that has long been a staple of work in the constructivist tradition. If individuals are motivated not by their narrow material welfare but by concerns for the well-being of those with whom they share a common identity, nationalism or other appeals may well produce a "national" interest that is distinctive from that deduced from aggregating individual interests.

No theoretical approach to international relations is immune from the challenges of the aggregation process. If the new behavioral revolution offers more robust insights into how individuals obtain and process information then similar tools can be applied to the question of groups. Such a research program would include experimental work on how groups with heterogeneous preferences operating in different institutional settings might generate different decisions. Such work could help address big, important questions such as how institutions exacerbate or dampen individual-level biases. Engaging the field of comparative politics, international relations scholars could also explore how different decision-making processes (e.g., parliamentary or presidential systems) or bureaucratic organizations (e.g., meritocratic or politicized) might differentially channel heterogeneous preferences rooted in behavioral research. Work of this sort has barely begun—including in this special issue—but this line of research promises fruitful arbitrage between the behavioral revolution and the study of international relations.

Two Cautions

While we believe that behavioral approaches hold promise for the study of international relations, there are two important methodological critiques of this emerging literature. The first critique concerns external validity. Most studies in this issue—and in the field more generally—use surveys or experiments to identify the effects of key psychological traits on behavior, preferences, beliefs, and decision making. A handful of these experiments and surveys—such as those by Bayram here—are conducted on decision-making elites, but they are more frequently run on random samples or even samples of convenience, most notably of college students.[83] Psychologists have long made peace with such sampling techniques, but they may be less appropriate for a field in which the object of study is elite decision makers and in which the stakes of actual decision making diverge radically from what can be replicated in the lab.

We have two cautious rejoinders to this first criticism. First, external validity is often portrayed as a methodological problem: can inferences be drawn from sample populations to elites? But if a theory is expected to apply to any decision

83. Other studies that rely on elite samples include Hafner-Burton et al. 2014; Hafner-Burton, LeVeck and Victor 2016a, 2016b; Renshon 2015; and Tomz 2008.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

maker, then testing that theory in a convenience sample, even one that may not be reflective of the broader population, is appropriate. The researcher needs to assure only that the "treatment" being investigated is randomly assigned across that sample, allowing an appropriate identification of any causal effect. The criticism would apply only if analysts do not state fully the scope conditions of their theory, such as why they would expect the psychological effects in question to operate on only certain classes of subjects. External validity is more of a problem, however, where elites are shown to exhibit very different characteristics than other populations, as we have noted they might.

Administering a survey on risk or discounting to twenty-year-old students and claiming to have found a general psychological bias is as problematic as administering such a survey to sophisticated financial analysts or traders and drawing general conclusions.[84] Similarly, if high stakes are expected to elicit more reflective System-2 thinking, testing a theory on students who receive only class credit or Mechanical Turk respondents who receive only a token payment would be equally problematic. But it is equally inappropriate to condemn all behavioral studies of foreign policy decision making because they may be based on convenience samples. This is particularly true if the theories under consideration are rooted in decision-making models that derive expectations from rational choice processes that are held similarly to pertain more generally.

The second critique of behavioral research involves replicability. Many of the insights that underpin the behavioral revolution are based on empirical findings; recent attempts at replication have raised questions about which of these are truly robust.[85] Movements within the field to improve quality control over experimental protocols and to allow publication of null results may help to address this broader concern.[86] But these problems are hardly unique to international relations work, and extend from psychology to behavioral economics and even medicine.

Conclusion

We have tried to capture the essence of the behavioral revolution by contrasting it with the foundations of rational choice. We noted areas where the two approaches are complementary—with new behavioral work introducing new twists on preferences, beliefs, and decision making that could be incorporated into rationalist models—and others, for example, with respect to emotion, in which more fundamental differences may be at work. We then sketched how this new work is connecting to

84. See Tetlock 2005 for an example.

85. Camerer et al. 2016; Open Science Collaboration 2015.

86. For this movement, which is dynamic and growing quickly, see the Berkeley Initiative for Transparency in the Social Sciences at http://www.bitss.org/, and Evidence in Governance and Politics (EGAP) at http://egap.org/.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

international relations, highlighting directions in which it is likely to evolve as it moves forward.

In contrast to the larger revolution, work in international relations appears to be distinctive in its emphasis on the effects of individual heterogeneity on choice processes and strategic interactions. This can be seen in areas of research as diverse as the effect of group preferences on the tendency to cooperate or how actors might respond to changes in the distribution of capabilities. The behavioral revolution in international relations also promises to shed new light on political and decision-making processes and even institutional design, cooperation, and compliance.

Thomas Kuhn famously argued that scholars bandwagon to a new paradigm to explain anomalies long before the underlying theory is completely explicated.[87] Whether the behavioral revolution is a new paradigm in Kuhn's sense of the term is very much an open question. Even in economics, where it has been discussed at greater length than in political science, the "revolution" is compatible with some of neoclassical theory, less compatible with other parts, and quite subversive in still others.

Although we expect similar processes to operate in international relations, there are still strongly differing opinions on the most promising ways forward. We have two contrasting conclusions. Stein argues that the new behavioral work is likely to have more traction within the discipline this time around than it did when psychological studies first surfaced in international relations in the 1970s and 1980s, in part because of its experimental foundation. Powell is more skeptical, arguing that the new behavioral work risks opening even greater distance between deductively valid theories and empirical research in international relations. Clearly, this special issue will not close the debate on how best to approach the study of international politics. We hope, however, that it will stimulate fruitful new research.

References

Akerlof, George A., and Janet L. Yellen. 1990. The Fair Wage-Effort Hypothesis and Unemployment. *Quarterly Journal of Economics* 105 (2):255–83.

Alevy, Jonathan E., Michael S. Haigh, and John A. List. 2007. Information Cascades: Evidence from a Field Experiment with Financial Market Professionals. *Journal of Finance* 62 (1):151–80.

Allais, Maurice. 1953. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica* 21 (4):503–46.

Allison, Graham T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. New York: Little, Brown.

Alpert, Marc, and Howard Raiffa. 1982. A Progress Report on the Training of Probability Assessors. In *Judgement Under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Amos Tversky, and Paul Slovic, 294–305. Cambridge, UK: Cambridge University Press.

87. Kuhn 1962.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Andreoni, James. 1995. Cooperation in Public Goods Games: Kindness or Confusion? *American Economic Review 85* (4):891–904.
Ardanaz, Martin, M. Victoria Murillo, and Pablo M. Pinto. 2013. Sensitivity to Issue Framing on Trade Policy Preferences: Evidence from a Survey Experiment. *International Organization* 67 (2):411–37.
Aumann, Robert. 1976. Agreeing to Disagree. *Annals of Statistics* 4 (6):1236–39.
Aumann, Robert, and Adam Brandenburger. 1995. Epistemic Conditions for Nash Equilibrium. *Econometrica* 63 (5):1161–80.
Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.
Banks, Jeffrey S. 1990. Equilibrium Behavior in Crisis Bargaining Games. *American Journal of Political Science* 34 (3):599–614.
Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. A Model of Investor Sentiment. *Journal of Financial Economics* 49 (3):307–43.
Bayram, A. Burcu. 2011. Collective Identity and International Law: The Individual Psychology of Compliance. Unpublished manuscript, University of Texas, Arlington, Texas.
———. 2016a. Good Europeans? How European Identity and Costs Interact to Explain Compliance with European Union Law. *Journal of European Public Policy*. Available at: <http://www.tandfonline.com/doi/abs/10.1080/13501763.2015.1114659>.
———. 2016b. Aiding Strangers: Generalized Trust and the Moral Basis of Public Support for Foreign Development. *Foreign Policy Analysis*. First published online 7 May at <http://fpa.oxfordjournals.org/content/early/2016/06/07/fpa.orw008>.
———. 2017. Due Deference: Cosmopolitan Social Identity and the Psychology of Legal obligation in International Politics. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S137–63.
Bechara, Antoine, Hanna Damasio, and Antonio R. Damasio. 2000. Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex* 10 (3):295–307.
Bendor, Jonathan, and Thomas H. Hammond. 1992. Rethinking Allison's Models. *American Political Science Review* 86 (2):301–22.
Berejikian, Jeffrey D. 2002. A Cognitive Theory of Deterrence. *Journal of Peace Research* 39 (2):165–83.
Biersteker, Thomas J., and Cynthia Weber, eds. 1996. *State Sovereignty as Social Construct*. Cambridge UK: Cambridge University Press.
Blainey, Geoffrey. 1973. *The Causes of War*. London: Macmillan.
Bleiker, Roland, and Emma Hutchison. 2008. Fear No More: Emotions and World Politics. *Review of International Studies* 34 (S1):115–35.
Broude, Tomer. 2015. Behavioral International Law. *University of Pennsylvania Law Review* 163 (4): 1099–157.
Burgoon, Brian A., and Michael J. Hiscox. 2004. The Mysterious Case of Female Protectionism: Gender Bias in Attitudes Toward International Trade. Paper presented at the Annual Meeting of the American Political Science Association, August, Philadelphia, PA.
Byman, Daniel L., and Kenneth M. Pollack. 2001. Let Us Now Praise Great Men: Bringing the Statesman Back In. *International Security* 25 (4):107–46.
Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2002. Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games. *Journal of Economic Theory* 104 (1):137–88.
———. 2003. Models of Thinking, Learning, and Teaching in Games. *American Economic Review* 93 (2): 192–95.
———. 2004. A Cognitive Hierarchy Model of Games. *Quarterly Journal of Economics* 119 (3):861–98.
Camerer, Colin F., George Loewenstein, and Drazen Prelec. 2005. Neuroeconomics: How Neuroscience Can Inform Economics. *Journal of Economic Literature* 43 (1):9–64.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. Evaluating Replicability of Laboratory Experiments in Economics. *Science* 351 (6280):1433–36.

Checkel, Jeffrey T. 1998. The Constructivist Turn in International Relations Theory. *World Politics* 50 (2): 324–48.

Chi, Michelene T.H., Robert Glaser, and Ernest Rees. 1982. Expertise in Problem Solving. In *Advances in the Psychology of Human Intelligence*, Vol. 1, edited by Robert J. Sternberg, 7–75. Hillsdale, NJ: Erlbaum.

Colgan, Jeff D. 2013. Domestic Revolutionary Leaders and International Conflict. *World Politics* 65 (4): 656–90.

Coller, Maribeth, and Melonie B. Williams. 1999. Eliciting Individual Discount Rates. *Experimental Economics* 2 (2):107–27.

Conlisk, John. 1996. Why Bounded Rationality? *Journal of Economic Literature* 34 (2):669–700.

Costa-Gomes, Miguel A., and Vincent P. Crawford. 2006. Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *American Economic Review* 96 (5):1737–68.

Crawford, Neta C. 2009. Human Nature and World Politics: Rethinking "Man." *International Relations* 23 (2):271–88.

Croco, Sarah E. 2011. The Decider's Dilemma: Leader Culpability, War Outcomes, and Domestic Punishment. *American Political Science Review* 105 (3):457–77.

Davis, James W. Jr. 2000. *Threats and Promises: The Pursuit of International Influence*. Baltimore, MD: Johns Hopkins University Press.

Debs, Alexandre, and H.E. Goemans. 2010. Regime Type, the Fate of Leaders, and War. *American Political Science Review* 104 (3):430–45.

Debs, Alexandre, and Nuno P. Monteiro. 2014. Known Unknowns: Power Shifts, Uncertainty, and War. *International Organization* 68 (1):1–31.

DellaVigna, Stefano. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature* 47 (2):315–72.

Edelstein, David M. 2002. Managing Uncertainty: Beliefs About Intentions and the Rise of Great Powers. *Security Studies* 12 (1):1–40.

El-Gamal, Mahmoud A., and David M. Grether. 1995. Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association* 90 (432):1137–45.

Engelmann, Dirk, and Martin Strobel. 2004. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review* 94 (4):857–69.

Farnham, Barbara. 1994. *Avoiding Losses/Taking Risks: Prospect Theory and International Conflict*. Ann Arbor, MI: University of Michigan Press.

Fearon, James D. 1994. Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88 (3):577–92.

———. 1998. Bargaining, Enforcement, and International Cooperation. *International Organization* 52 (2): 269–305.

Fearon, James D., and Alexander Wendt. 2002. Rationalism v. Constructivism: A Skeptical View. In *Handbook of International Relations*, edited by Walter Carlsnaes, Thomas Risse, and Beth Simmons, 52–72. Thousand Oaks, CA: Sage.

Fehr, Ernst, and Colin F. Camerer. 2007. Social Neuroeconomics: The Neural Circuitry of Social Preferences. *Trends in Cognitive Sciences* 11 (10):419–427.

Fehr, Ernst, and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114 (3):817–68.

Fehr, Ernst, and Simon Gächter, 2000. Cooperation and Punishment in Public Goods Experiments. *The American Economic Review* 90 (4):980–94.

Fey, Mark, and Kristopher W. Ramsay. 2011. Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict. *American Journal of Political Science* 55 (1):149–69.

Finnemore, Martha. 1996. *National Interests in International Society*. Ithaca, NY: Cornell University Press.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Finnemore, Martha, and Kathryn Sikkink. 1998. International Norm Dynamics and Political Change. *International Organization* 52 (4):887–917.
Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. Are People Conditionally Cooperative? Evidence From a Public Goods Experiment. *Economics Letters* 71 (3):397–404.
Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits. 2015. The Distributional Preferences of an Elite. *Science* 349 (6254), doi:10.1126/science.aab0096.
Fowler, James H., and Cindy D. Kam. 2006. Patience as a Political Virtue: Delayed Gratification and Turnout. *Political Behavior* 28 (2):113–28.
———. 2007. Beyond the Self: Social Identity, Altruism, and Political Participation. *Journal of Politics* 69 (3):813–27.
Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature* 40 (2):351–401.
Frieden, Jeffry A. 1999. Actors and Preferences in International Relations. In *Strategic Choice and International Relations*, edited by David A. Lake and Robert Powell, 39–76. Princeton, NJ: Princeton University Press.
Frieden, Jeffry A., and Ronald Rogowski. 1996. The Impact of the International Economy on National Policies: An Analytical Overview. In *Internationalization and Domestic Politics*, edited by Robert O. Keohane and Helen V. Milner, 25–47. New York: Cambridge University Press.
Galbraith, Jean. 2013. Treaty Options: Towards a Behavioral Understanding of Treaty Design. *Virginia Journal of International Law* 53 (2):309–63.
Geanakoplos, John. 1992. Common Knowledge. *Journal of Economic Perspectives* 6 (4):53–82.
Goldstein, Judith, Yotam Margalit, and Douglas Rivers. 2008. Producer, Consumer, Family Member: The Relationship Between Trade Attitudes and Family Status. Paper prepared for the Conference on Domestic Preferences and Foreign Economic Policy, Princeton University, April, Princeton, NJ.
Gourevitch, Peter A. 1999. The Governance Problem in International Relations. In *Strategic Choice and International Relations*, edited by David A. Lake and Robert Powell, 137–64. Princeton, NJ: Princeton University Press.
Gries, Peter Hays. 2004. *China's New Nationalism*. Berkeley, CA: University of California Press.
Guisinger, Alexandra. 2014. Racial Diversity and Redistribution: Explaining (White) Americans' Continued Support for Trade Protection. Paper presented at the Annual Meeting of the American Political Science Association, August, Washington DC.
Hafner-Burton, Emilie M., D. Alex Hughes, and David G. Victor. 2013. The Cognitive Revolution and the Political Psychology of Elite Decision Making. *Perspectives on Politics* 11 (2):368–86.
Hafner-Burton, Emilie M., Brad L. LeVeck, David G. Victor. 2016a. How Activists Perceive the Utility of International Law. *Journal of Politics* 78 (1):167–80.
———. 2016b. No False Promises: How the Prospect of Non-Compliance Affects Elite Preferences for International Cooperation. *International Studies Quarterly* doi:10.1093/isq/sqw047
Hafner-Burton, Emilie M., Brad L. LeVeck, David G. Victor, and James H. Fowler. 2014. Decision Maker Preferences for International Legal Cooperation. *International Organization* 68 (4):845–76.
Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Vintage Books.
Haigh, Michael S., and John A. List. 2005. Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis. *Journal of Finance* 60 (1):523–34.
Halevy, Yoram. 2007. Ellsberg Revisited: An Experimental Study. *Econometrica* 75 (2):503–36.
Hall, Rodney B. 1999. *National Collective Identity: Social Constructs and International Systems*. New York: Columbia University Press.
Hall, Todd H., and Andrew A.G. Ross. 2015. Affective Politics after 9/11. *International Organization* 69 (4):847–79.
Hall, Todd, and Keren Yarhi-Milo. 2012. The Personal Touch: Leaders' Impressions, Costly Signaling, and Assessments of Sincerity in International Affairs. *International Studies Quarterly* 56 (3):560–73.
Harsanyi, John C. 1967. Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model. *Management Science* 14 (3):159–82.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Harvey, Frank P. 2012. *Explaining the Iraq War: Counterfactual Theory, Logic, and Evidence*. New York: Cambridge University Press.
Hemmer, Christopher M. 2000. *Which Lessons Matter? American Foreign Policy Decision-Making in the Middle East, 1979–1987*. Albany, NY: State University of New York Press.
Herrmann, Richard K. 2009. *Perceptions and Behavior in Soviet Foreign Policy*. Pittsburgh, PN: University of Pittsburgh Press.
———. 2017. How Attachments to the Nation Shape Beliefs About the World: A Theory of Motivated Reasoning. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S61–S84.
Herrmann, Richard K., and Michael P. Fischerkeller. 1995. Beyond the Enemy Image and Spiral Model: Cognitive–Strategic Research after the Cold War. *International Organization* 49 (3):415–50.
Hiscox, Michael J. 2002. *International Trade and Political Conflict: Commerce, Coalitions, and Mobility*. Princeton, NJ: Princeton University Press.
———. 2006. Through a Glass and Darkly: Attitudes Toward International Trade and the Curious Effects of Issue Framing. *International Organization* 60 (3):755–80.
Ho, Daniel E., and Kosuke Imai. 2008. Estimating Causal Effects of Ballot Order from a Randomized Natural Experiment: The California Alphabet Lottery, 1978–2002. *Public Opinion Quarterly* 72 (2): 216–40.
Horowitz, Michael, Rose McDermott, and Allan C. Stam. 2005. Leader Age, Regime Type, and Violent International Relations. *Journal of Conflict Resolution* 49 (5):661–85.
Horowitz, Michael, Allan Stam, and Cali M. Ellis. 2015. *Why Leaders Fight*. New York: Cambridge University Press.
Houghton, David Patrick. 1998. Historical Analogies and the Cognitive Dimension of Domestic Policymaking. *Political Psychology* 19 (2):279–303.
Huberman, Gur. 2001. Familiarity Breeds Investment. *Review of Financial Studies* 14 (3):659–80.
Hymans, Jacques E.C. 2006. *The Psychology of Nuclear Proliferation: Identity, Emotions and Foreign Policy*. Cambridge, MA: Cambridge University Press.
Ikenberry, G. John. 2001. *After Victory: Institutions, Strategic Restraint, and the Rebuilding of Order after Major Wars*. Princeton, NJ: Princeton University Press.
Janis, Irving L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Boston, MA: Houghton Mifflin.
Jervis, Robert. 1970. *The Logic of Images in International Relations*. Princeton, NJ: Princeton University Press.
———. 1976. *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.
———. 1992. Political Implications of Loss Aversion. *Political Psychology* 13 (2):187–204.
Johnson, Dominic D.P., and James H. Fowler. 2011. The Evolution of Overconfidence. *Nature* 477 (7364): 317–20.
Jolls, Christine, Cass R. Sunstein, and Richard Thaler. 1998. A Behavioral Approach to Law and Economics. *Stanford Law Review* 50 (5):1471–550.
Jupille, Joseph, Walter Mattli, and Duncan Snidal. 2013. *Institutional Choice and Global Commerce*. New York: Cambridge University Press.
Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.
Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2):263–91.
Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.
Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. Experimental Tests of the Endowment Effect and the Coase Theorem. *Journal of Political Economy* 98 (6):1325–48.
Katzenstein, Peter J., ed. 1996. *The Culture of National Security: Norms and Identity in World Politics*. New York: Columbia University Press.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Kertzer, Joshua D. 2017. Resolve, Time, and Risk. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S109–S136.
Kertzer, Joshua D., and Kathleen M. McGraw. 2012. Folk Realism: Testing the Microfoundations of Realism in Ordinary Citizens. *International Studies Quarterly* 56 (2):245–58.
Khong, Yuen Foong. 1992. *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965*. Princeton, NJ: Princeton University Press.
Klotz, Audie. 1996. Norms and Sanctions: Lessons from the Socialization of South Africa. *Review of International Studies* 22 (2):173–90.
Krasner, Stephen D. 1976. State Power and the Structure of International Trade. *World Politics* 28 (3): 317–47.
———, ed. 1983. *International Regimes*. Ithaca, NY: Cornell University Press.
Kratochwil, Friedrich V. 1989. *Rules, Norms, and Decisions: On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs*. Cambridge, UK: Cambridge University Press.
Krebs, Ronald R., and Aaron Rapport. 2012. International Relations and the Psychology of Time Horizons. *International Studies Quarterly* 56 (3):530–43.
Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Kydd, Andrew H. 2010. Learning Together, Growing Apart: Global Warming, Energy Policy and International Trust. *Energy Policy* 38 (6):2675–80.
Laibson, David. 1997. Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics* 112 (2): 443–78.
Lake, David A. 2009. Open Economy Politics: A Critical Review. *Review of International Organization* 4 (3):219–44.
———. 2010–11. Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War. *International Security* 35 (3):7–52.
Lake, David A., and Robert Powell. 1999. *Strategic Choice and International Relations*. Princeton, NJ: Princeton University Press.
Larson, Deborah Welch. 1989. *Origins of Containment: A Psychological Explanation*. Princeton, NJ: Princeton University Press.
Ledyard, John. 1995. Public Goods: A Survey of Experimental Research. In *The Handbook of Experimental Economics*, edited by John Kagel and Alvin E. Roth, 111–94. Princeton, NJ: Princeton University Press.
LeVeck, Brad, D. Alex Hughes, James H. Fowler, Emilie Hafner-Burton, and David G. Victor. 2014. The Role of Self-interest in Elite Bargaining. *Proceedings of the National Academy of Sciences of the United States of America* 111 (52):18536–41.
Levy, Jack S. 1992a. An Introduction to Prospect Theory. *Political Psychology* 13 (2):171–86.
———. 1992b. Prospect Theory and International Relations: Theoretical Applications and Analytical Problems. *Political Psychology* 13 (2):283–310.
———. 1997. Prospect Theory, Rational Choice, and International Relations. *International Studies Quarterly* 41 (1):87–112.
Lieberman, Matthew D., Darren Schreiber, and Kevin N. Ochsner. 2003. Is Political Cognition Like Riding a Bicycle? How Cognitive Neuroscience Can Inform Research on Political Thinking. *Political Psychology* 24 (4):681–704.
Loewenstein, George, and Jon Elster, eds. 1992. *Choice Over Time*. New York: Russell Sage Foundation.
Loomes, Graham, Chris Starmer, and Robert Sugden. 1991. Observing Violations of Transitivity by Experimental Methods. *Econometrica* 59 (2):425–39.
Lopez, Anthony C., Rose McDermott, and Michael Bang Petersen. 2011. States in Mind: Evolution, Coalitional Psychology, and International Politics. *International Security* 36 (2):48–83.
Lu, Xiaobo, Kenneth Scheve, and Matthew J. Slaughter. 2012. Inequity Aversion and the International Distribution of Trade Protection. *American Journal of Political Science* 56 (3):638–54.
Mansfield, Edward D., and Diana C. Mutz. 2009. Support for Free Trade: Self-Interest, Sociotropic Politics, and Out-Group Anxiety. *International Organization* 63 (3):425–57.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Massey, Cade, and Richard H. Thaler. 2010. The Loser's Curse: Decision Making and Market Efficiency in the National Football League Draft. Working paper.
McDermott, Rose. 1998. *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy*. Ann Arbor, MI: University of Michigan Press.
———. 2007. *Presidential Leadership, Illness, and Decision Making*. New York: Cambridge University Press.
McNamara, Kathleen. 1999. *The Currency of Ideas: Monetary Politics in the European Union*. Ithaca, NY: Cornell University Press.
Mercer, Jonathan. 2005. Rationality and Psychology in International Politics. *International Organization* 59 (1):77–106.
———. 2010. Emotional Beliefs. *International Organization* 64 (1):1–31.
Moravcsik, Andrew. 1997. Taking Preferences Seriously: A Liberal Theory of International Politics. *International Organization* 51 (4):513–53.
Nagel, Rosemarie. 1995. Unraveling in Guessing Games: An Experimental Study. *American Economic Review* 85 (5):1313–26.
Naoi, Megumi, and Ikuo Kume. 2011. Explaining Mass Support for Agricultural Protectionism: Evidence from a Survey Experiment During the Global Recession. *International Organization* 65 (4):771–95.
Neuman, Russell, George E. Marcus, Michael Mackuen, and Ann N. Crigler, eds. 2007. *The Affect Effect: Dynamics of Emotion in Political Thinking and Behavior*. Chicago: University of Chicago Press.
Neustadt, Richard E., and Ernest R. May. 1988. *Thinking in Time: The Uses of History for Decision Makers*. New York: Free Press.
O'Donoghue, Ted, and Matthew Rabin. 1999. Doing It Now or Later. *American Economic Review* 89 (1): 103–24.
Open Science Collaboration. 2015. Estimating the Reproducibility of Psychological Science. *Science* 349 (6251):943–53.
Pauwelyn, Joost. 2015. The Rule of Law Without the Rule of Lawyers? Why Investment Arbitrators Are from Mars, Trade Adjudicators from Venus. *American Journal of International Law* 109 (4):761–805.
Plott, Charles R., and Kathryn Zeiler. 2005. The Willingness to Pay–Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* 95 (3):530–45.
Poulsen, Lauge. 2013. When the Claim Hits: Bilateral Investment Treaties and Bounded Rational Learning. *World Politics* 65 (2):273–313.
Powell, Robert. 2017. Research Bets and Behavioral IR. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S265–S277.
Prather, Lauren. 2014. Values at the Water's Edge: Social Welfare Values and Foreign Aid. Unpublished manuscript, Stanford University, Stanford, CA
Rabin, Matthew. 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83 (5):1281–302.
Rathbun, Brian C. 2009. It Takes All Types: Social Psychology, Trust, and the International Relations Paradigm in Our Minds. *International Theory* 1 (3):345–80.
———. 2011. Before Hegemony: Generalized Trust and the Creation and Design of International Security Organizations. *International Organization* 65 (2):243–73.
Rathbun, Brian C., Joshua D. Kertzer, and Mark Paradis. 2017. *Homo Diplomaticus*: Mixed-Method Evidence of Variation in Strategic Rationality. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S33–S59.
Reiter, Dan. 1996. *Crucible of Beliefs: Learning, Alliances, and World Wars*. Ithaca, NY: Cornell University Press.
Renshon, Jonathan. 2015. Losing Face and Sinking Costs: Experimental Evidence on the Judgment of Political and Military Leaders. *International Organization* 69 (3):659–95.
Renshon, Jonathan, Julia J. Lee, and Dustin Tingley. 2015. Physiological Arousal and Political Beliefs. *Political Psychology* 36 (5):569–85.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

———. 2017. Emotions and the Micro-Foundations of Commitment Problems. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S189–S217.
Rho, Sungmin, and Michael Tomz. 2017. Why Don't Trade Preferences Reflect Economic Self-Interest? In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1): S85–S108.
Rogowski, Ronald. 1999. Institutions as Constraints on Strategic Choice. In *Strategic Choice and International Relations*, edited by David A. Lake and Robert Powell, 115–36. Princeton, NJ: Princeton University Press.
Rubinstein, Ariel. 1989. The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge." *American Economic Review* 79 (3):385–91.
———. 1998. *Modeling Bounded Rationality*. Cambridge, MA: MIT Press.
Sabet, Shahrzad. 2014. Feelings First: Non-Material Factors as Moderators of Economic Self-Interest Effects on Trade Preferences. Unpublished manuscript, Princeton University, Princeton, NJ
Saunders, Elizabeth N. 2009. Transformative Choices: Leaders and the Origins of Intervention Strategy. *International Security* 34 (2):119–61.
———. 2017. No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S219–S247.
Schwarz, Norbert. 2000. Emotion, Cognition, and Decision Making. *Cognition and Emotion* 14 (4):433–40.
Shafer, D. Michael. 1988. *Deadly Paradigms: The Failure of US Counterinsurgency Policy*. Princeton, NJ: Princeton University Press.
Simon, Herbert A. 1955. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics* 69 (1): 99–118.
Sitaraman, Ganesh, and David Zionts. 2015. Behavioral War Powers. *New York University Law Review* 90: 516–88.
Stahl, Dale O., and Paul W. Wilson. 1994. Experimental Evidence on Players' Models of Other Players. *Journal of Economic Behavior and Organization* 25 (3):309–27.
———. 1995. On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior* 10 (1):218–54.
Stein, Janice Gross. 1985. Calculation, Miscalculation, and Conventional Deterrence, II: The View from Jerusalem. In *Psychology and Deterrence*, edited by Robert Jervis, Richard Ned Lebow, and Janice Gross Stein, 60–88. Baltimore, MD: Johns Hopkins University Press.
Stein, Janice Gross, and Louis W. Pauly. 1992. Choosing to Co-Operate: How States Avoid Loss. *International Journal* 47 (2):199–201.
Steinbruner, John D. 1977. *The Cybernetic Theory of Decision: New Dimensions of Political Analysis*. Princeton, NJ: Princeton University Press.
Strotz, R.H. 1955. Myopia and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies* 23 (3):165–80.
Suskind, Ron. 2007. *The One Percent Doctrine: Deep Inside America's Pursuit of its Enemies Since 9/11*. New York: Simon and Schuster.
Svenson, Ola. 1981. Are We All Less Risky and More Skillful than Our Fellow Drivers? *Acta Psychologica* 47 (2):143–48.
Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.
Tetlock, Philipe E., and Daniel Gardner. 2015. *Superforecasting: The Art and Science of Prediction*. New York: Crown.
Thaler, Richard H. 1980. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization* 1 (1):39–60.
———. 1999. Mental Accounting Matters. *Journal of Behavioral Decision Making* 12 (3):183–206.
Tingley, Dustin. 2017. Rising Power on the Mind. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S165–S217.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press

Tomz, Michael. 2007. Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization* 61 (4):821–40.

———. 2008. Reputation and the Effect of International Law on Preferences and Beliefs. Working paper, Stanford University, Stanford, CA

Tversky, Amos, and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185 (4157):1124–32.

———. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5 (4):297–323.

Van Aaken, Anne. 2014. Behavioral International Law and Economics. *Harvard International Law Journal* 55 (2):421–81.

Wagner, R. Harrison. 2000. Bargaining and War. *American Journal of Political Science* 44 (3):469–84.

Weeks, Jessica L. 2008. Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization* 62 (1):35–64.

———. 2012. Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict. *American Political Science Review* 106 (2):326–47.

Wendt, Alexander. 1999. *Social Theory of International Relations*. New York: Cambridge University Press.

Wilson, Rick K. 2011. The Contribution of Behavioral Economics to Political Science. *Annual Review of Political Science* 14 (1):201–23.

Yoshida, Wako, Ben Seymour, Karl J. Friston, and Raymond J. Dolan. 2010. Neural Mechanisms of Belief Inference During Cooperative Games. *Journal of Neuroscience* 30 (32):10744–51.

Zimbardo, Philip G., and John N. Boyd. 1999. Putting Time in Perspective: A Valid, Reliable Individual-Differences Metric. *Journal of Personality and Social Psychology* 77 (6):1271–88.

https://doi.org/10.1017/S0020818316000400 Published online by Cambridge University Press