Research Bets and Behavioral IR

Robert Powell

Abstract Behavioral IR faces a fundamental challenge. The actors in most IR models and theories are not individuals—they are aggregates like states, ministries, interest groups, political parties, and rebel factions. There are two broad approaches to attempting to integrate behavioral research about individuals. The first, a quasi-behavioral approach, makes nonstandard assumptions about the preferences, beliefs, or decision-making processes of aggregate actors. The second tries to build theories in which the key actors are individuals. Pursuing the former means that the assumptions about actors will be only weakly linked to the empirical findings propelling behavioral research. The second approach faces formidable obstacles that international relations theory has confronted for a long time and for the most part has not overcome.

Behavioral economics has been one of the most active and productive areas of research in economics in recent years. Given the close methodological parallels between economics and rational-choice approaches to international relations (IR) theory, it is natural to ask how and to what extent the insights, ideas, findings, and, more broadly, the approach of behavioral economics will affect IR theory. The essays in this special issue try to provide a "framework for integrating the new behavioral revolution into international relations."[1] The issue "aims to chart a strategy for incorporating the new behavioral revolution more fully into international relations."[2] I offer some conjectures about the likely impact of the behavioral revolution on the study of international relations and highlight some of the challenges a truly behavioral approach to international relations theory faces. The ultimate impact, of course, will not be known for some time and will depend on the "research bets" scholars make and which of them pay off.[3] Watching those research bets play out in the coming years will be very exciting.

The compelling empirical evidence driving behavioral economics is that individuals deviate in systematic ways from the predictions of the standard model. However, the actors in most IR models are not individuals—they are aggregates like states, ministries, interest groups, political parties, rebel groups, etc. Incorporating the empirical findings in IR models thus poses a significant challenge.

There are two broad approaches to dealing with the mismatch between behavioral findings about individuals and the fact that the actors in most IR models and theories are aggregate actors. One approach, which might be called quasi-behavioral, is to

1. Hafner-Burton et al. 2017, S4.
2. Ibid., S3.
3. See Lake and Powell 1999 for a discussion of research bets in international relations.

International Organization 71, Supplement 2017, pp. S265–S277
© The IO Foundation, 2017.
doi:10.1017/S0020818316000461

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

continue to assume that groups can be adequately treated as unitary actors but that these unitary actors deviate from the standard rational-actor assumptions in the same ways that behavioral research has shown individuals deviate. For example, states, like individuals, have social or time-inconsistent preferences. A second, more direct approach is to try to build theories in which the key actors are individuals.

The first way of dealing with the mismatch means that the assumptions about actors will be only weakly linked to the empirical findings propelling behavioral research. The compelling empirical evidence that individuals deviate in systematic ways from the standard assumptions is not compelling evidence for or against treating groups as unitary actors with behavioral or standard preferences. Treating a group of individuals as if it was a unitary actor with well-defined preferences, whether behavioral or standard, is a huge assumption that we make for pragmatic reasons having to do with tractability and other considerations. Without a compelling account of how individual preferences and actions aggregate up into a group outcome, treating groups as unitary actors with standard preferences or behavioral preferences is a pragmatic "research bet" about which assumption will lead to interesting findings and insights. Indeed, the biggest bet may be that we can treat the group as a unitary actor with well-defined preferences of any kind. Once we have made this bigger bet, we should be open to either standard or behavior assumptions and be guided by our intuition about which is more likely to be fruitful.

The second approach to trying to build theories in which the key actors are individuals faces formidable obstacles that international relations theory has confronted for a long time. Behavioral findings may have sparked renewed interest in building bottom-up theories in which the actors actually are individuals. Indeed, the essays in this issue generally focus on individuals and try to varying degrees to connect the preferences of individuals to "international" outcomes. But even if "work of this sort has barely begun" as Hafner-Burton and colleagues observe, it is not clear why the renewed interest should make the old obstacles any less formidable.[4] Bottom-up theories are very, very hard to construct.[5]

Behavioral economics builds on economics, and economic models generally place a great deal of emphasis on establishing a tight connection between (some) assumptions and the conclusions and predictions that follow from them. This tight theoretical linkage has been crucial to the success of behavioral economics. It makes it possible to show that there are systematic deviations from the predictions of the standard model and that different assumptions like social preferences or time-inconsistent preferences can account for some of these deviations. By contrast, the link between assumptions and conclusions is much weaker in much of international relations theory, and this is likely to limit the effect the behavioral approach has on IR theory.

If we avoid this limitation by focusing on tight deductive models like game-theoretic formalizations, then we face another kind of limitation. The tools we have to

4. Hafner-Burton et al. 2017, S21.
5. Stein shares the view that this kind of theory is very hard to develop. Stein 2017.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

forge tight links are themselves quite weak and limited. Models quickly become too complicated to analyze if one tries to do too much. If a tight deductive framework is a *sine qua non*, then tractability considerations may force us to choose between investing in better models of individuals or better models of the strategic environment in which actors interact. Is the greatest expected return on our research bets to be found in better assumptions about individuals or in better assumptions about the strategic environment?

The Rational Choice Paradigm in Economics

Behavioral economics builds on and is able to leverage a well-established paradigm. It is worth quoting Matthew Rabin, a leading behavioral economist, at some length:

> Mainstream economics employs a powerful combination of methods: methodological individualism, mathematical formalization of assumptions, logical analysis of the consequences of those assumptions, and sophisticated empirical field testing.

But, he goes on,

> "tractability" and "parsimony" should be guiding principles to make our research more realistic, not pretexts for avoiding this task. As it now stands, some important psychological findings seem tractable and parsimonious enough that we should begin the process of integrating them into economics. Incorporating other findings will take longer. But even in those cases, economists ought to become aware of the shortcomings of our models, regret these shortcomings, and keep our eyes open for ways to remedy them.[6]

The fulcrum of the behavioral lever is a tight link between a clear set of assumptions and the conclusions that follow from them. This link is usually formal, but the essential feature is that it is tight. This makes two things possible. First, actual outcomes can be compared to specific predictions and shown to deviate from these predictions in significant and systematic ways. This comparison would be much harder and much less compelling if the connection between assumptions and conclusions were looser. Second, the implications of alternative assumptions can be derived and shown to account for some or all of these deviations.

A fundamental difference between economics and IR is that the link between assumptions and conclusions in many IR theories is weaker than in economics. This difference makes it harder to trace the effects of more empirically or

6. Rabin 1998, 12–13.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

psychologically grounded assumptions. This suggests that the way that behavioral economics has developed may be a very poor guide to how behavioral IR will develop.

Bargaining with Social Preferences

Having a concrete example of a behavioral IR model will help to clarify and focus the discussion. That said, there is nothing new here. Indeed, one of the main points of the analysis is that substituting behavioral preferences for standard preferences is conceptually pretty pedestrian stuff once we have a well-defined model.

Suppose that two states, S_1 and S_2, are bargaining over a "pie" of size 1. We can think of the pie as the benefits the states get from reaching a negotiated settlement. Let x_j be j's share where $j \in \{1, 2\}$. S_1 gets a payoff $U_1(x_1, x_2)$ and S_2 gets $U_2(x_1, x_2)$ from the settlement (x_1, x_2).[7]

Following Fearon and Powell, assume S_1 makes a take-it-or-leave-it offer to S_2 who can accept or fight. If S_2 accepts, S_j gets $U_j(x_1, x_2)$.[8] If S_2 fights, S_j wins everything with probability p_j (where $p_1 + p_2 = 1$) and pays cost c_j. Finally, suppose that S_1 is uncertain of S_2's cost but believes it to be uniformly distributed over $[\widehat{c}_2 - u, \widehat{c}_2 + u]$ where $\widehat{c}_2$ is S_2's average cost and u is a measure of S_1's uncertainty about S_2's cost.

A standard assumption about preferences is that an actor cares only about its share, for example, S_1's utility depends on only x_1. If so, we can write $U_1(x_1, x_2)$ as $U_1(x_1)$. Assuming risk neutrality then gives $U_1(x_1) = x_1$.

The quasi-behavioral approach makes behavioral assumptions about aggregate actors who in this case are states. Assume more specifically that the states have social preferences. Each state's payoff depends on its payoff *and* the other state's payoff. To keep things simples, let $U_1(x_1, x_2) = (1 - \lambda_1)x_1 + \lambda_1\ x_2 = x_1 + \lambda_1\ (x_2 - x_1)$ and similarly for S_2. The parameter λ_1 measures how much S_1 cares about S_2's share x_2. If $\lambda_1 = 0$, we have the standard assumption that S_1 cares about only its own absolute share x_1. S_1 is altruistic if it gains when the other state gains ($\lambda_1 > 0$). S_1 exhibits "relative-gains concerns" when $\lambda_1 < 0$. That is, S_1's payoff is decreasing in the gap between S_2's payoff and its own. In keeping with the old debate in international relations theory about relative gains concerns, we will focus on the case of relative gains concerns ($\lambda_1 < 0$) and let $r_1 = -\lambda_2$ to ease the notation. This leaves $U_1(x_1, x_2) = x_1 - r_1(x_2 - x_1)$. The larger r_1, the more severe S_1's concerns about relative loses.

Before analyzing the model, it may be useful to point out the rationalist versus not rationalist tension highlighted in the introduction seems to play little if any role here.

7. It is convenient to describe a settlement explicitly in terms of each state's share. If the states have to fully allocate the pie and cannot agree to an inefficient outcome, then $x_1 + x_2 = 1$ and an outcome can be characterized in terms of either state's share via $x_2 = 1 - x_1$.

8. Fearon 1995; Powell 1999.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

A very common goal in formalization is to try to have the standard assumptions appear as a special case of more general formulation as happens here when the concern for others is 0 ($\lambda_j = r_j = 0$). A more general formulation makes it possible to trace the effects of behavioral preferences (subject to tractablity issues).

The more S_1 offers, the more likely S_2 is to accept and the less likely war is. But the more S_1 offers, the less it will have if S_2 accepts. S_1's optimal offer balances these competing considerations and is

$$x_2^* = p_2 + \frac{c_1}{2(1+2r_1)} - \frac{\widehat{c}_2 + u}{2(1+2r_2)}.$$

This offer will generally be rejected with positive probability and result in an equilibrium probability of war w^* equal to

$$w^* = \frac{3}{4} - \frac{\widehat{c}_2}{4u} - \frac{c_1(1+2r_2)}{4u(1+2r_1)}$$

(see the appendix).

At least in this setting, it is easy to trace the effects of the states' social preferences on the outcome. More specifically, the effect of relative gains concerns depends on the state or, rather, whether it is the state making the offer or responding to it. The more S_1 is concerned about relative gains (the higher r_1), the less it offers and the higher the risk of war (x_2^* decreases and w^* increases). By contrast, the more S_2 is concerned about relative gains, the more S_1 offers and the lower the risk of war (x_2^* is increasing in r_2 and w^* is decreasing). If both states are equally concerned about relative gains (i.e., if $r_1 = r_2$), then these concerns have no effect on the probability of cooperation. The probability of cooperation (i.e., avoiding war) is $1 - w^* = (u + \widehat{c}_2 + c_1)/(4u)$.

Two Approaches to Behavioral IR

The actors in the bargaining model are not individuals. S_1 and S_2 are states. Indeed, the actors in most IR models and theories are not individuals. They are aggregates that are assumed to behave as if they were unitary actors with well-defined preferences. Thus there is a certain mismatch between the compelling empirical evidence about individual behavior that drives behavioral economics and the actors typically assumed in IR models and theories.

Broadly speaking, there are two ways of dealing with this mismatch and of trying to incorporate behavioral insights and findings into international relations. The analysis of the bargaining model illustrates the first. The quasi-behavioral approach assumes that aggregate actors like states can continue to be treated as unitary actors but now with nonstandard preferences and nonstandard beliefs like projection

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

bias or overconfidence. It also assumes that they may engage in nonstandard decision making.[9]

At the risk of belaboring the point, a prerequisite for this approach is a tight deductive link between assumptions and conclusions as there is, for example, in the model I describe. Without that, it is hard to see how one can make much progress in understanding what follows from what and in particular what follows from assuming behavioral preferences rather than standard preferences.

More importantly, the main justification for this quasi-behavioral approach to dealing with the mismatch is pragmatic. Namely, a researcher bets that treating groups as unitary actors but now with behavioral preferences will strike a fruitful balance between tractability and descriptive accuracy and ultimately yield interesting insights and results. The compelling empirical evidence about individuals may motivate this bet, but this evidence does not in fact support nonstandard assumptions about group behavior. Whether individuals with nonstandard preferences in a particular organizational structure aggregate up into a group that has well defined preferences—be they standard or nonstandard—is an open question that most analyses avoid by assumption in the name of tractability. Indeed, as Hafner-Burton and colleagues observe, organizations may mitigate the effects of nonstandard individual behavior.[10] Institutions may, for example, help overcome commitment problems resulting from time-inconsistent individual preferences.[11]

Time and future research will ultimately determine how fruitful this way of dealing with the mismatch is. But the bargaining model and the absolute-and-relative gains debate in IR are suggestive. Indeed, this debate is illuminating in part because we now have some historical perspective on it. Recall that Kenneth Waltz and neorealists more generally argued that states care about relative gains and that this concern significantly limits international cooperation.[12] Robert Keohane challenged this conclusion in *After Hegemony*, arguing that neorealism's core assumptions do not imply little or no cooperation.[13] More specifically, he claimed that the repeated prisoner's dilemma was consistent with neorealism's core assumptions and that, moreover, cooperation was possible in this game. Joseph Grieco countered that the repeated prisoner's dilemma was not consistent with neorealism's core assumptions.[14] The states in Keohane's repeated prisoner's dilemma were only concerned with their (long-run) absolute gains. By contrast, realism

9. DellaVigna 2009 categorizes the work in behavioral economics in this way and provides an excellent survey of it. We will sometimes ease the exposition by using "nonstandard preferences" or "behavioral preferences" as a shorthand to refer to all three of these categories.
10. Hafner-Burton et al. 2017.
11. Also see Saunders's contribution to this volume on this point.
12. Waltz 1979.
13. Keohane 1983.
14. Grieco 1988.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

> expects a state's utility function to incorporate *two distinct terms.* It needs to include the state's individual payoff V, reflecting the realist view that states are motivated by absolute gains. Yet it must also include a term integrating both the state's individual payoff and the partner's payoff, W, in such a way that gains favoring the state add to this utility while, more importantly, gaps favoring the partner detract from it. One function that depicts this realist understanding of state utility is $U = V - k(W - V)$, with k representing the state's coefficient of sensitivity to gaps in payoffs either to its advantage or disadvantage.[15]

In earlier work I contested Grieco's claim that realism expects a relative-gains utility function.[16] Engaging in a bit of exegesis, I argued that a better reading of Waltz and neorealism was that a state's concerns for relative gains is induced by the structure of the strategic environment and that this is not well modeled by representing this concern in states' utility functions. More importantly, I presented a model in which the environment induces the concern for relative gains in actors that are solely concerned with their long-run absolute gains.[17]

What should one take away from this brief review of the absolute-and-relative gains debate which, in retrospect, may have been one of the first quasi-behavioral efforts in IR? The compelling empirical evidence that individual behavior is often inconsistent with the predictions of the standard model is pretty much irrelevant to whether we can usefully treat the state as a unitary actor with standard or behavioral preferences. Rather, the key question in the debate was whether seeing relative gains concerns as an attribute of the state or as something induced by the structure of the strategic environment would prove more insightful and empirically fruitful. Would it be more useful to assume nonstandard preferences, that is, a concern for relative gains, was impeding cooperation? Or would it be more useful to assume standard preferences and then strive to understand how structural features like the intensity of the security dilemma could induce these concerns?[18] I am hardly a disinterested observer of this debate, but it appears to me at least that the latter approach to understanding relative gains concerns clearly proved to be the more illuminating.

Economics and international relations are obviously very different fields. It is nevertheless useful to note that economists have not pursued a quasi-behavioral approach. Economists have typically not dealt with the mismatch between evidence about individual behavior and models with aggregate actors by making nonstandard assumptions about the aggregate actors. Firms, for example, are assumed to have standard profit-maximizing preferences, or the deviation from profit maximization

15. Keohane 1983, 500. This formulation of relative gains concerns is, of course, also what was assumed above in the bargaining model where r_1 rather than k measured S_1's concern about relative gains and in which $U_1(x_1, x_2) = x_1 - r_1 (x_2 - x_1)$.
16. Powell 1991, 1994.
17. Powell 1991.
18. See Glaser 1994–95.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

is something to be explained, perhaps by agency issues within the firm, rather than something to be assumed.[19]

The reluctance of behavioral economists to make nonstandard assumptions about aggregate actors may be traceable to the heavy emphasis they put on empirical evidence as the key motivation for making nonstandard assumptions. One can clearly show in the controlled environment of the laboratory that the standard model of expected utility maximization cannot explain how individuals often act. These empirical findings are the rejoinder to the normal appeals to tractability and parsimony as the long quotation from Rabin makes clear. A key question for behavioral economists thus becomes: how do better and more empirically realistic assumptions about individuals aggregate up to affect market or firm behavior or more broadly social outcomes? Are deviations from the standard model "large enough to matter for our theories of how markets and institutions work?"[20] One cannot get at this by just assuming that a group of individuals can be treated as a unitary actor with well-defined nonstandard preferences.

The second way of dealing with the mismatch between empirical evidence about individuals and aggregate actors is to take it on directly. There are two variants to this direct approach. The first starts with individuals and actually shows that their preferences along with the organizational structure of the entity combine to produce something that behaves as if it were a unitary actor with nonstandard preferences. The parallel here is with the Median Voter Theorem and related results. Under a very limited set of circumstances a group of voters with standard preferences behaves as if it were a rational unitary actor whose preferences mirror that of the median voter.[21] A natural conjecture is that a group of individuals with nonstandard preferences will also behave as if it were a unitary actor with well-defined nonstandard preferences in a very limited set of circumstances. In light of this, a more promising variant of a head-on approach is to start with individuals, bypass the intermediate stage of trying to show that groups behave as if they were unitary actors with well-defined preferences, and attempt instead to link these individuals to international outcomes like war or trade policy. In effect, the direct approach would be a first-image explanation in which individuals' nonstandard preferences could be deductively linked to international outcomes. It is parallel to efforts in economics to see how behavioral assumptions affect market behavior.

This is the general approach taken by the papers in this special issue. For example, Rho and Tomz focus on survey responses of individuals (voters).[22] Kertzer's experimental study of the sources of resolve has undergraduates playing a bargaining game in a laboratory.[23] Tingley conducts a field experiment on shifting power via a

19. This of course begs the question of how the standard assumptions became standard and therefore priviledged. See Tirole 1988 for a discussion of the profit-maximization hypothesis.

20. DellaVigna 2009, 316.

21. More specifically, this result holds when the group makes decisions by majority voting on a one-dimensional issue, each voter has symmetric, single-peaked preferences, and there is a unique median.

22. Rho and Tomz 2017.

23. Kertzer 2017.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

survey.[24] Interesting and thought provoking as the findings about individuals are, they are not directly connected to international outcomes. Of course, there is no reason that these particular papers must directly connect these findings about individuals to international outcomes. But the discipline as a whole must if the compelling empirical evidence about individuals is to be directly linked to the outcomes of international politics.

Formulating this kind of first-image IR model or theory would be a tremendous achievement. Hafner-Burton and colleagues may also be correct in suggesting that behavioral findings will reopen "the aggregation issue in international relations."[25] But it is hard to be optimistic about the prospects of such a research program. Starting with individuals with nonstandard properties may not be any more difficult than starting with individuals with standard properties. The latter however has so far proved beyond our grasp when it comes to many questions in international relations. Instead, researchers have generally pursued what seem like more fruitful paths. They make pragmatic compromises by taking a boxes-within-boxes approach and dealing at a level of aggregation that seems appropriate to the problem at hand.[26]

The First Wave

DellaVigna, Rabin, and other prominent behavioral economists prefer the appellation "psychology and economics" to "behavioral economics." Economics in general is about explaining behavior. The distinctive feature of their endeavor is integrated psychological findings into economic theory. Thinking of "psychology and IR" rather than "behavioral IR" helps make it clear that the work highlighted in this issue and the research program outlined in the introduction is really the second effort to integrate psychology into international relations theory. Robert Jervis, Irving Janis, Richard Ned Lebow, and Janice Stein all made major contributions in the 1970s and 1980s to what might be called the first wave of behavioral IR.[27] What does the first wave of behavioral IR suggest about the likely future of the second?

The first wave was full of insights. Its portrayal of individual actors' misperceptions and biases was certainly more descriptively accurate than that of the rational-actor approach. The first-wave work got the most traction when the questions of interest were directly related to individuals, for example, how individual leaders assess other leaders or are vulnerable to motivated biases.

But the first wave, like IR theory more generally, had a harder time linking individual characteristics to many international outcomes like the causes of war.

24. Tingley 2017.
25. Hafner-Burton et al. 2017, S5.
26. See Lake and Powell 1999, 13–16 and Hafner-Burton et al. 2017 for discussions of the boxes-within-boxes approach.
27. See, for example, Jervis's (1970, 1976) path-breaking work on misperception; Janis 1972 on group think; and Jervis, Lebow, and Stein 1985 on psychology and deterrence. Jack Levy (1992) brought prospect theory into the picture somewhat later. Levy 2013, Goldgeiger and Tetlock 2001, and Stein 2017 (this issue) discuss this first wave of work.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

Confronted with this challenge, the first wave often opted for the quasi-behavioral approach. Namely, one assumed that aggregate units exhibit the same kinds of misperceptions that the individuals in it do. For example, in Jervis's discussion of self-fulfilling prophesies and the spiral model, the "crucial question is the degree to which a state's actions that are based on an initially false image have transformed the other state's intentions."[28] The aggregate actor of the state is assumed here to have intentions and initially false images. Similarly, work on prospect theory assumed that states as well as individuals exhibit framing effects.

The pragmatic assumption that states rather than individuals misperceive or (more generally) are uncertain of another state's type has proved very fruitful. Among other things, it is the basis for modern theories of costly signaling and escalation. However, this pragmatic approach is at best only weakly linked to the psychological work on misperception that centers on individuals. The first wave of behavioral IR never got around the aggregation problem as the lack of first-image theories suggests. This experience may indicate that the second wave will also find it very difficult to get around this problem.

Research Bets in Behavioral IR

Behavioral IR can be seen as an effort to relax some of the very strong simplified assumptions we make about actors' preferences and decision making. In some cases like the bargaining model analyzed above, it is very easy to trace the effects of these efforts. But there is likely to be a trade-off in other more complex settings between relaxing assumptions about actors and about, say, the strategic environment in which they act. As noted at the outset, a fundamentally important prerequisite to understanding the implications of any set of assumptions is a tight deductive link between assumptions and conclusions. Unfortunately, the tools we have to forge these links are themselves weak and quite limited. Tight deductive models, like game-theoretic formalizations, quickly become too complicated and unwieldy if one tries to do too much. If a tight deductive framework is a *sine quo non*, then the tractability considerations Rabin mentions may force us to choose between investing in models with nonstandard and often more complicated assumptions about actors' preferences or in models with richer and more complicated assumptions about the strategic environment. If we face this kind of trade-off, is the greatest expected return on our research bets to be found in making nonstandard assumptions about preferences or in richer assumptions about the actors' environment?

There is of course no way to know whether investing in more elaborate models of preferences or the strategic environment offers the highest expected payoff. That is why it is a research bet. And, indeed, the discipline is probably better off if different scholars make different bets. But as the relative-gains debate suggests, what often

28. Jervis 1976, 77.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

seems puzzling from a rational choice perspective only appears that way because of quite strong (and sometimes underarticulated) assumptions about the environment in which the actors interact. A better description of the strategic environment (which may include disaggregating actors into more basic aggregate actors), along with the standard assumptions, has frequently gone a very long way to resolving the puzzle.

Caveats

A fundamental mismatch bedevils efforts to develop an empirically based behavioral theory of international relations. The compelling empirical evidence is that individual behavior deviates in systematic ways from the standard model. But the actors in IR models are rarely individuals and are much more typically groups that are assumed to be unitary actors with well-behaved preferences.

There are two general ways of dealing with this mismatch. A quasi-behavioral approach assumes that unitary actors have behavioral preferences. The advantage of this approach is that different assumptions about, say, states' preferences can be readily incorporated in existing frameworks. The primary disadvantage is that the empirical evidence about individual preferences is not evidence about group preferences. Whatever a researcher assumes about a group's unitary preferences, he or she is making a pragmatic judgment about theoretical constructs and in particular about which assumptions about a fictitious unitary actor are likely to provide useful insights.

The second, more direct way of dealing with the mismatch is to eliminate it by developing theories in which individuals are the primary actors. The obvious advantage to this approach is that the behavioral evidence about individuals is much more directly relevant and should be accorded more weight when the actors in the theory actually are individuals. The obvious disadvantage is that bottom-up, first-image theories have so far proved very difficult to construct.

In sum, the effect of behavioral IR on the field as a whole is likely to remain quite modest as long as individuals play a small role in our best theoretical and empirical efforts to answer the IR questions we care about. In their introductory essay, Hafner-Burton and colleagues ask, "How broad of a challenge does the behavioral revolution pose to rationalist models?"[29]. The emphasis of this question seems misplaced to me. A more apt question may be: How broad of a challenge does the behavioral revolution pose to approaches based on treating groups individuals as unitary actors? Whether we assume a unitary actor has behavorial or standard preferences is a second-order assumption compared to assuming groups to be unitary actors with well-defined preferences.

29. Hafner-Burton et al. 2017, S3.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

Appendix

Analyzing the model when the states have social preferences is straightforward. S_j's payoff to fighting with relative-gains concerns is $F_1 = p_1 U_1(1,\ 0) + (1 - p_1)U_1(0,\ 1) - c_1 = (1 - p_2)(1 + r_1) + p_2(-r_1) - c_1 = 1 + r_1 - p_2(1 + 2r_1) - c_1$. S_2's payoff is $F_2 = -r_2 + p_2(1 + 2r_2) - c_2$. If S_1 offers $(x_{1,}\ x_2)$ type c_2 accepts if $U_2(x_1,\ x_2) \geq F_2$ or, equivalently, if $(1 + r_2)x_2 - r_2(1 - x_2) \geq -r_2 + p_2(1 + 2r_2) - c_2$ (where we assume nothing is left on the table so that $x_1 + x_2 = 1$). Simplifying, c_2 accepts if $c_2 \geq (p_2 - x_2)(1 + 2r_2)$. This implies that the probability that S_2 accepts x_2 is $[\widehat{c}_2 + u - (1 + 2r_2)(p_2 - x_2)]/(2u)$. S_1's expected payoff to offering x_2 is

$$V_1(x_2) = [(1 + r_1)(1 - x_2) - r_1 x_2]\left(\frac{\widehat{c}_2 + u - (p_2 - x_2)(1 + 2r_2)}{2u}\right) + [1 + r_1 - p_2(1 + 2r_1) - c_1]\left(\frac{(p_2 - x_2)(1 + 2r_2) - \widehat{c}_2 + u}{2u}\right).$$

Solving the first-order condition for the optimal offer gives

$$x_2^* = p_2 + \frac{c_1}{2(1 + 2r_1)} - \frac{\widehat{c}_2 + u}{2(1 + 2r_2)}.$$

This yields an equilibrium probability of war of

$$w^* = \frac{3}{4} - \frac{1}{4u}\left(\widehat{c}_2 - \frac{c_1(1 + 2r_2)}{1 + 2r_1}\right).$$

If the states are equally concerned with relative gains ($r_1 = r_2 \equiv r$), then the optimal offer and probability of war are

$$x_2^* = p_2 - \frac{\widehat{c}_2 + u - c_1}{2(1 + 2r)}$$

and

$$w^* = \frac{3}{4} - \frac{\widehat{c}_2 + c_1}{4u}.$$

References

DellaVigna, Stefano. 2009. Psychology and Economics. *Journal of Economic Literature* 47 (2):315–72.
Fearon, James. 1995. Rationalist Explanations for War. *International Organization* 49 (3):379–414.
Glaser, Charles. 1994–95. Realists As Optimists: Cooperation As Self-Help. *International Security* 19 (3): 50–90.
Goldgeiger, James, and Philip Tetlock. 2001. Psychology and International Relations Theory. *Annual Review of Political Science* 4:67–92.
Grieco, Joseph. 1988. Anarchy and the Limits of Cooperation. *International Organization* 42 (3):485–507.
Hafner-Burton, Emilie, Stephan Haggard, David A. Lake, and David G. Victor. 2017. The Behavioral Revolution and International Relations. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S1–S31.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press

Janis, Irving. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decsisions 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.

Jervis, Robert, Richard Ned Lebow, and Janice Gross Stein, eds. 1985. *Psychology and Deterrence*. Baltimore, MD: Johns Hopkins University Press.

Keohane, Robert. 1983. *After Hegemony*. Princeton, NJ: Princeton University Press.

Kertzer, Joshua D. 2017. Resolve, Time, and Risk. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S109–S136.

Lake, David, and Robert Powell. 1999. International Relations: A Strategic-Choice Approach. In *Strategic Choice and International Relations*, edited by David Lake and Robert Powell, 3–38. Princeton, NJ: Princeton University Press.

Levy, Jack. 1992. Prospect Theory and International Relations: Theoretical Applications and Analytical Problems. *Political Psychology* 13 (2):283–310.

———. 2013. Psychology and Foreign Policy Decision-Making. In *The Oxford Handbook of Political Psychology*, 2nd ed. edited by Leonie Huddy, David O. Sears, and Jack S. Levy, 301–33. New York: Oxford University Press.

Niou, Emerson, and Peter Ordeshook. Stability in Anarchical International Systems. *American Political Science Review* 84 (4):1207–34.

Powell, Robert. 1991. Absolute and Relative Gains in International Relations Theory. *American Political Science Review* 85 (4):1303–20.

———. 1994. Anarchy in International Relations Theory. *International Organization* 48 (2):313–44.

———. 1999. *In the Shadow of Power*. Princeton, NJ: Princeton University Press.

Rabin, Matthew. 1998. Psychology and Economics. *Journal of Economic Literature* 36 (1):11–46.

Rho, Sungmin, and Tomz Michael. 2017. Why Don't Trade Preferences Reflect Economic Self-Interest? In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S85–S108.

Saunders, Elizabeth. 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.

Stein, Janis Gross. 2017. The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S249–S263.

Tingly, Dustin. 2017. Rising Power on the Mind. In The Behavioral Revolution and International Relations, supplement, *International Organization* 71 (S1):S165–S188.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Waltz, Kenneth. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.

https://doi.org/10.1017/S0020818316000461 Published online by Cambridge University Press