Who's Afraid of "Rational Choice Theory"?

R. Harrison Wagner*

October, 2001

> A spectre is haunting Europe—the spectre of communism. All the powers of old Europe have entered into a holy alliance to exorcise this spectre: Pope and Tsar, Metternich and Guizot, French Radicals and German police-spies.
>
> It is high time that Communists should openly, in the face of the whole world, publish their views, their aims, their tendencies, and meet this nursery tale of the spectre of communism with a manifesto of the party itself.
>
> *The Communist Manifesto*

With apologies to Marx and Engels, the specter that is haunting political science these days is not communism, but something that is commonly called "rational choice theory," and many political scientists are busily engaged in forming coalitions among otherwise disparate factions to oppose it. One casualty of their opposition has been a proper understanding of what this specter actually consists of. Another, I fear, may be the training of graduate students whose only interest is in acquiring the skills they will need for successful careers in political science. It is for them that I have decided to confront this nursery tale of the specter of rational choice theory with the following personal manifesto.

Theorize this!

It is impossible to understand what is at stake in this controversy if one does not understand what a valid argument is, why the validity of arguments is important, and why it is often difficult to construct valid arguments or to determine whether an argument is valid or not. The significance of valid arguments can be illustrated in the simplest possible way by a visit to the dog pound.

Suppose a man wants to find an inexpensive dog for his children to play with, and goes to the dog pound to look for one. He naturally wants a dog that will be good with his children, and not one that would maul them. Suppose the attendant assures him that a particular dog would have that quality. Skeptical, he might ask, "How do you know that, and why should I believe it?" The attendant might reply that the dog in question was a Labrador retriever. "So?" the man might reply, to which the attendant

*Professor, Department of Government, University of Texas, Austin, TX 78712 USA. Email: rhwagner@mail.utexas.edu.

might respond that Labrador retrievers are good with children.

This man has been given perhaps the simplest possible argument that actually conveys new information. It has two premises: "Labrador retrievers are good with children" and "This dog is a Labrador retriever," from which it follows that "This dog will be good with children," which is what he wanted to know. The conclusion "follows from" the premises only because if one accepts the premises and denies the conclusion one would have contradicted oneself, which is why if one believes the premises one must also believe the conclusion. Arguments that have this property are called valid arguments, and reasoning from premises to conclusion in this way is commonly called "deductive reasoning."

However, this little argument would satisfy such a man only if he were confident that both the premises were true. If one or both were not, the argument would remain valid but the conclusion might be false. Suppose, then, that the man asked why he should believe that this dog was a Labrador retriever—this is, after all, the dog pound. The attendant might reply that Labrador retrievers had certain recognizable characteristics such as a large, square head, short hair, a wide chest, and a friendly disposition, and this dog had those characteristics.

At first glance this looks like a deductive argument just like the first one: the first premise is that Labrador retrievers have certain recognizable characteristics, and the second is that this dog has all those characteristics. But if so, the argument is not valid because it does not follow from these two premises that the dog is a Labrador retriever. Such an argument would be an example of a logical fallacy called "affirming the consequent," and therefore could not provide the assurance the man was looking for.

But this would be a misunderstanding of the atten-

dant's reasoning. The attendant is saying that the hypothesis that the dog is a Labrador retriever would explain its appearance, and thus its appearance gives us reason to believe that it is a Labrador retriever. There is a deductive argument here, but its premises are, "All Labrador retrievers have certain recognizable characteristics," and "This dog is a Labrador retriever," from which it would follow, if true, that this dog would have the properties of a Labrador retriever. But this is something that one does not have to be persuaded of, since the dog can be inspected directly. The question is, rather, what sort of dog is it? And the reasoning is that since these premises, if true, would imply that the dog would have the appearance that it does have, the fact that it has that appearance is evidence that the premises are true. This is an example of what is commonly called "inductive reasoning," and the problem of induction is to figure out what justifies an inference of this sort.

Probability theory provides a plausible answer to that question, since such an inference can be shown to be an application of Bayes' rule. (Howson and Urbach 1993) But whether that provides an adequate account of inductive reasoning or not is irrelevant: we do not require a justification for reasoning in this way to do it, and the notion that some philosopher might discover that we are mistaken in doing so is preposterous. What is important here is, rather, the fact that the inductive inference from the dog's visible characteristics to its breed is made possible by a deductive inference from the breed to a dog's visible characteristics: if the breed could not explain its appearance, then the breed could not be inferred from the appearance.[1] The problem is that there are other possible explanations of the dog's appearance, some of which might imply that it would be dangerous, and that is why inductive inference requires not just identifying a possible explanation of the facts, but also supplying reasons to believe that that explanation is better than other possible ones. Thus inductive inference is sometimes said to be "inference to the best explanation." (Harman 1965, Lipton 1991)

Similarly, if the man asked why he should believe that Labrador retrievers are good with children, he might be told that many people had had such dogs as pets and this was their uniform experience. Since the hypothesis that all Labrador retrievers are good with children would explain the fact that everyone who had had them as pets found them to be good with children, that fact is evidence that the hypothesis is true. However, a cautious person might wonder if there were other possible explanations of this fact.

These examples illustrate a simple point: whether we are reasoning from premises to conclusions, or from observable facts to possible explanations of those facts, what is commonly called logical validity is necessary if our reasoning is to affect our beliefs: if the confidence we place in some premises is to be transferred to a conclusion then the conclusion must be implied by the premises, and if some explanation is to be supported by the facts then the facts must be implied by the explanation.

To see the relevance of all this to the question at hand, consider one of the most influential books about international politics ever written, *Theory of International Politics*, by Kenneth Waltz. Waltz claims in this book that international politics is characterized by the recurrent formation of balances of power. He offers an explanation of this fact in the following passage:

> If states wished to maximize power, they would join the stronger [of two states or coalitions], and we would see not balances forming but a world hegemony forged. This does not happen because balancing, not bandwagoning, is the behavior induced by the system. The first concern of states is not to maximize power but to maintain their positions in the system.
>
> Secondary states, if they are free to choose, flock to the weaker side; for it is the stronger side that threatens them. (Waltz 1979, 126–127)

Here Waltz is clearly saying that the fact that weak states band together against strong ones (thereby forming balances of power) cannot be explained if states want to maximize their power, but can be explained if they want instead to maximize their chances of survival. The premises that constitute the explanation are, "The first concern of states is to maintain their positions in the system," and "The stronger one's ally, the bigger a threat it could pose to one's position in the aftermath of victory." It is hard to doubt the second premise, but the fact that balances of power recur is the reason Waltz gives for believing the first one. Thus Waltz is claiming that since these premises, if true, would explain the observed fact, the observed fact is evidence that the first premise is true.

However, unlike our example from the dog pound, in this case the alleged fact to be explained does not follow from the premises, since it is also obviously true that the weaker one's ally, the less likely it is that one will be victorious. Thus in choosing allies a state must consider both the probability of victory in war and the danger its ally would pose in the aftermath of war if they are victorious. Knowing only that the first concern of states is to maintain their positions in the system is not enough to tell us what choice states should be expected to make.[2]

---

[1] In the Bayesian interpretation of inductive inference, the deductive argument tells us that the conditional probability of the conclusion being true, given the truth of the premises, is one.

[2] For recent discussions of this issue, see Powell 1999, chapter 5; and Wagner 2001. One could also question, of course, whether the fact that a strong state *could* be a threat to one's position should be a concern if there was reason to believe that it would not be. But Waltz's argument is

Some critics of Waltz's book have argued that to evaluate his argument it must be tested empirically, and to do that it is necessary to determine if alliances really do lead to balances of power. But Waltz claims not just that balances of power occur, but also that this fact can be explained in the way just discussed. Even if we could establish that alliances did (always? often? sometimes?) lead to balances of power, this could not possibly tell us that Waltz's explanation of that fact was correct, since a non sequitur cannot be used to explain anything. And even if we could establish that balances of power never occurred, we could not infer that states were not primarily concerned about "maintaining their positions in the system," since we do not know what they would do if they were. Moreover, if we cannot figure out what to expect if states are solely concerned about maintaining their positions, it is not clear how we could figure out what to expect if they were concerned about other things as well. Thus if we are to hope to explain anything we cannot avoid the construction of valid arguments such that the fact or facts to be explained can be derived from the premises that constitute the explanation.

Critics of what is commonly called "rational choice theory" typically either do not understand this requirement, or underestimate the difficulty involved in satisfying it. One has written, for example:

> Rather than engage in method-driven political science, it makes more sense to start with a problem: why do politicians who lose elections sometimes attempt coups rather than accept the popular verdict? Why do democracies seem better able to prevent famines than nondemocracies? What leads some people to vote for ethnic parties rather than those that appeal to economic interest or some other ideology, and why does this vary from time to time and place to place? Once the question is stated, the search should begin for the most viable explanation.... The goal should be to get the right answer, not to vindicate a pet approach. The most promising way to advance toward it is to develop empirical generalizations that can be tested by the predictions they make, modified when they fail, and tested again and again and again. Perhaps this inductive approach will add up to a general theory of politics one day, perhaps not. (Shapiro 2000)

"Perhaps" does not enter into it: empirical generalizations are not explanations, they are facts to be explained, and no accumulation of facts can produce an explanation of anything. That requires a creative guess as to what premises the fact or facts to be explained can be derived from, and then, if necessary, a demonstration that they can in fact be explained in that way.

---

not valid even if one assumes that it would be.

Apparently Shapiro believes that we should be satisfied if we could establish that, as Waltz claimed, balances of power often occur, and not trouble ourselves with trying to explain that fact. But without an explanation we cannot even be sure that a generalization based on this fact is true, since our observations may be an artifact of something we have overlooked, or chance. Moreover, we can never be confident that we might have overlooked some other possible generalization, perhaps, for example, by failing to investigate a possible correlation between war and relative shoe sizes. Finally, in international politics, at any rate, there are good reasons to believe that one should not expect to find many interesting empirical regularities, and therefore to try to understand international politics by looking for them may be like trying to uncover the laws of physics by looking for patterns in the weather.[3]

Stephen Walt, on the other hand, claims that it is easy to construct explanations, but he confuses logical validity with logical consistency and claims that it is consistency that we should aim for. (Walt 1999) A consistent argument is just one that is not self-contradictory, while a valid argument is one that is "so constructed that if the premises are jointly asserted the conclusion cannot be denied without a contradiction" (*Random House Dictionary of the English Language*, 1987). The problem with Waltz's argument is not that it is self-contradictory but that it is not valid. You should try for yourself to see how easy it would be to do better.[4]

Waltz claimed both that "balancing" was a general characteristic of international politics, and that he had explained it. But incomplete explanations in the political science literature are not restricted to attempts to establish general truths, they are also found in attempts to explain individual events. To take one example at random, consider the problem of explaining the occurrence of the US Civil War. To do so, it is necessary to explain (1) the decision by the states that joined the Confederacy to secede, (2) the decision by the remaining states to resist the secession of the southern states with force, and (3) the inability of the two sides to avoid war by reaching a compromise settlement of their differences. Let us consider what Richard Bensel (1990) has to say about how to explain the second of these facts.

One possibility would be to follow Shapiro's suggestion and rely on empirical generalizations as a way of ex-

---

[3]Not even the elementary premises employed by the dog pound attendant in the example given above are simple generalizations. They are implied by more general propositions about the heritability of various traits of dogs, which are in turn supported by the experience of animal breeders. (Darwin's fundamental insight was that what animal breeders did could be accomplished without an animal breeder. For a reconstruction of his argument, see Mayr 1991, 72.)

[4]Munck makes the same mistake. (Munck 2001, 201) Walt even claims that since scientists regularly commit the fallacy of affirming the consequent, logical rigor can't be all that important for scientific work! (Walt 1999, 16)

plaining what happened. Bensel says that:

> . . . all modern states aside from monarchies have repressed the secession of contiguous territory and, in the absence of foreign intervention, they have done so successfully in most cases. From this perspective, it would have been surprising, indeed, had the North chosen to let the South go without a fight.

However, he rightly rejects this as an adequate answer, since "the observation that all states repress separatism is an observation without a theoretical explanation. . . ." (Bensel 1990, 60)

Instead he offers the following explanation:

> The North suppressed secession because the separation of the South would have weakened, perhaps irremediably, bonds between the remaining states of the Union. In addition, loss of the South would have created a new political economy in which the basis of the Republican-led alliance of eastern industry and western yeoman agriculture would have rapidly dissolved as the two wings of the party struggled over the competing interests of industrial expansion and agricultural settlement. Thus, both national and party cohesion required suppression. (Bensel 1990, 93)

This is very interesting, but it is also very unclear how a decision by the northern states to suppress southern secession follows from the possibility that if the factions that made up the Republican Party had not united to fight the South they might have fallen out with each other afterward. Bensel has given us something to think about, but he has not given us an explanation of the fact he claimed to have explained. Nor would his explanation, even if it were complete, tell us why "all modern states aside from monarchies have repressed the secession of contiguous territory."

The political science literature is full of arguments and explanations such as these. Their prevalence is the result of two facts: (1) most writers do not recognize the importance of logical validity in constructing explanations, and (2) validity is often a very difficult criterion to satisfy. If the "pathology of rational choice theory" (Green and Shapiro 1994) is the belief that the conclusion of any complex valid argument must be true, the pathology of everything else is the belief that validity can be dispensed with. Sometimes people say that politics is just "not logical." But logic is not a property of the world, it is a property of what we say about the world. The world is a messy and confusing place. We do not enhance our understanding of it by saying messy and confusing things about it.[5]

[5]Walt points out that the reasoning in a number of important works has been found to be flawed, but they were nonetheless important. (Walt 1999, 16-17) This is no doubt true, but it is hardly a justification for

Beautiful models

The passages from Waltz and Bensel quoted above are examples of attempts to explain events by explaining the choices that led to them, and the clear assumption seems to be that these choices are to be explained by the preferences of the persons making them. This is what "rational choice" explanations are all about, and Waltz even claimed to be emulating microeconomics. Yet explanations such as these are never criticized by writers attacking "rational choice theory." Why not? The reason is that their complaint is not about rational choice explanations, flawed versions of which can be found everywhere, but rather about the use of models as a way of constructing them. And this complaint rests on a misunderstanding of what models are for and therefore how to evaluate them.[6]

A model in this context is just something that is used to represent something else, like a model airplane. Everyone who has used a map or a house plan or an architect's drawing has used a model. The purpose of such models is to facilitate inferences about the thing that is modeled that would otherwise be difficult. You could try to figure out how to landscape your yard or arrange the furniture in your new house just by standing in the middle of it and thinking about how it will look, but you might find it easier to work with a drawing. Similarly, you could give your guests complicated verbal instructions about how to find your house, but it might be more effective to give them a map and let them draw the proper inferences from it.

Whenever we use models such as these we have to worry whether conclusions that we reach that are true of the model also apply to the thing or things that the model represents. If a drawing of one's house or lawn is not drawn exactly to scale, then things that fit in the drawing won't fit in "the real world," and if roads that look straight on a map are really very crooked then it may take longer to get to your house than your friends thought. There are always differences of this sort between models and the things they represent, and the question therefore is not whether the model is completely accurate (no model is or can be, or it would not be a model), but whether it is accurate enough for the purpose at hand. A map that is good enough to enable people to find your house might not be good enough to determine how much fiber optic cable to buy if a company plans to wire your neighborhood, or to

accepting flaws in someone's else's reasoning, or for not trying to avoid them in one's own.

[6]An excellent introduction to the use of models in the social sciences can be found in Lave and March 1993. The notion that using people's preferences to explain their choices is restricted to "rational choice theory," or that the importance of doing that is in some way controversial or open to question, is the source of much confusion, for which both supporters and opponents of "rational choice theory" must share responsibility. For an example of a misleading discussion of "rational choice theory" by a supporter, see Bates 1997.

plot the path of a cruise missile.

The same is true of models of non-physical things. Formal or symbolic logic, for example, is a system of arbitrary symbols and rules for manipulating them that was designed to represent logical inference. Since the rules for manipulating the symbols are absolutely clear, it is often easier to prove theorems by using them than it is by using words. However, that led to controversies about whether theorems that are true in this symbolic language always carry over to the ordinary language that everyone actually thinks in. (Strawson 1952)

Similarly, what is called "rational choice theory" is really just a way of constructing models of people's choices, so it is not surprising that similar issues arise in that context. But to understand those issues we must appreciate the fact that they apply to the use of models of any sort in any context, and not just to rational choice models.

Since reasoning about models instead of the real thing can be misleading, there has to be a good reason for doing it. We should begin, then, by considering why people find a need to construct models of choices at all. Since we explain people's choices all the time without constructing models of them, the whole idea may seem ridiculous.[7]

There are three main ways in which explanations involving human choices can become complex enough that models of them can be useful. One is that the consequences of the choices of many people taken together may be nonobvious, and may then interact with people's subsequent choices. This is what happens in markets, and in electoral systems with competing political parties. A second is that individuals may be faced with uncertainty about the consequences of their choices, so their choices are not implied in any straightforward way by their preferences over final outcomes. And a third is that individuals' choices may be interdependent, in that what one person will choose depends on his or her expectations about how one or more other people will choose, and vice versa.

To see how the second and third of these possibilities can make things complicated, let us consider some controversies about how to explain what states did during the Cold War, using the Cuban Missile Crisis as an example. Graham Allison, in a book that was nearly as influential as Waltz's *Theory of International Politics*, discussed three different ways of explaining states' foreign policy decisions, one of which was to think of states as rational actors.[8] "The rational agent," he wrote, "selects the

alternative whose consequences rank highest in terms of his goals and objectives," and therefore rational actor explanations are characterized by the following "dominant inference pattern":

> If a nation performed a particular action, that nation must have had ends toward which the action constituted a maximizing means. The Rational Actor Model's explanatory power stems from this inference pattern. The puzzle is solved by finding the purposive pattern within which the occurrence can be located as a value-maximizing means. (Allison 1971, 33)

In illustrating this "inference pattern" with the Cuban Missile Crisis, Allison listed several alternative possible explanations of Khrushchev's decision to order the placement of ballistic missiles in Cuba. These explanations are distinguished simply by the objectives that Khrushchev hoped to accomplish by this means. One possibility, for example, is that he wanted to set up a deal whereby the US would remove its missiles from Turkey in exchange for the removal of the Soviet missiles in Cuba. Another is that he wanted to pressure the US to agree to a concession affecting the status of West Berlin, a third that he wanted to deter a possible US attack on Cuba, and so forth.

But this "inference pattern" cannot explain Khrushchev's decision to put missiles in Cuba. Khrushchev could have wanted (and no doubt did want) to accomplish all these objectives, but since putting missiles in Cuba could have had bad consequences for him rather than good ones, he might well have chosen not to put them there. And therefore his decision to put them there is not implied by any of these objectives alone. Putting missiles in Cuba was not like deciding which size wrench to take out of the toolbox, and the decision to do it cannot be explained in the way one might explain why a mechanic chose a large wrench rather than a small one. Thus there is actually no example of a "rational actor" explanation of foreign policy decisions to be found anywhere in Allison's famous book.

The fact that Kennedy had warned Khrushchev not to put missiles in Cuba led Richard Ned Lebow to argue that his decision to put them there anyway could not possibly be explained as the consequence of a rational choice. Rather, he claimed, it was an example of the general tendency of foreign policy makers to ignore information about the likely consequences of their actions, and therefore to be surprised by them, a tendency which could only be explained as the result of cognitive errors caused by stress and anxiety. (Lebow 1981)

Here, then, we appear to have two competing explanations of a famous foreign policy decision, and one might think that we must find a way to decide which is better. But in fact neither Allison nor Lebow has actually offered

---

[7]Why did the chicken cross the road? Because the expected utility of being on the other side was greater. (Readers puzzled by this footnote will have to look elsewhere for an explanation of chicken crossing the road jokes.)

[8]Allison 1971, 10–66. The two other types of explanations that Allison discussed focused on the effects of routine organizational operating procedures, and on intragovernmental or bureaucratic politics. Allison called these types of explanations "models," but they are really just summaries of the sort of reasoning to be found in three different bodies of

literature.

a possible explanation of what happened. As we have seen, Allison's explanation is a non sequitur and therefore cannot explain anything. Lebow's explanation, on the other hand, does not actually imply that we should have expected Khrushchev to do what he did either. It merely says that Khrushchev's decision was a mistake, which people subjected to various kinds of stress are likely to make. Thus without knowing what a rational decision would have been, Lebow's explanation has no content. But we don't know what a "rational choice" explanation would have led us to expect, and Lebow does not tell us. He merely asserts that had Khrushchev been acting rationally he would not have done what he did. What, then, are we to do?[9]

Clearly if we are to use Khrushchev's preferences to explain his actions we must somehow take into account what he believed the consequences of his actions would be—this is what is missing in Allison's account. And the question raised by Lebow's discussion is how uncertain Khrushchev could have been about them, since if he thought that a bad outcome was sufficiently unlikely he might have chosen to do what he did even if he were thinking clearly about what he should do. Lebow claims that in this and in most other crises, the state initiating the crisis had enough warning of an adverse response by its adversary that it should have been virtually certain of the response, and therefore its action can only be explained by the hypothesis that it ignored information that was readily available. How might we evaluate this claim?

Let us see how far we can get in answering this question without constructing a model. We know that Kennedy was in fact very reluctant to use military force in response to the missiles, and that if he had been unwilling to do so then nothing could have prevented Khrushchev from deploying them. Thus careful consideration could have left Khrushchev uncertain about how Kennedy would respond to them. Second, we know that Kennedy might have warned Khrushchev not to put missiles in Cuba even if he had been unwilling to use force if he did, since if Khrushchev had not done so then Kennedy could have accomplished his objective without using force. Now let us suppose, third, that it was common knowledge to Kennedy and Khrushchev that each would choose an action that led to his most preferred outcome, and fourth, that the answer we give to this question would be understood by them as well. These four premises imply that Lebow cannot be right in saying that Kennedy's warning should have been enough to make Khrushchev certain of a military response to the missiles. For suppose, to the contrary, that Kennedy's warning should be expected to convince Khrushchev not to put missiles in Cuba. Then Kennedy

would issue such a warning even if he were bluffing. But then the warning could convey no information to Khrushchev, and he would be foolish to take it seriously, which implies a contradiction.

Does this argument look like a "universal deductive theory" to you? Like an attempt to imitate physics? This is how critics of "rational choice theory" have characterized the specter they are trying to exorcise. (Shapiro 2000) But this argument is merely a verbal statement of the reasoning one finds in game-theoretic models.

But if we can reason like this without a model, what do we need a model for? Part of the answer is that the important question raised by Lebow's discussion of crises is not whether warnings by a leader of a state can make its adversary certain of how it will respond to a challenge, but whether such warnings can convey any information at all. That is a harder question to answer.

What would be required for the answer to be yes? Would you agree that it would be necessary that Khrushchev would believe that Kennedy would be more likely to issue such a warning if he were not bluffing than if he were? If so, then you might be reassured to learn that this is a straightforward implication of Bayes' rule, which game-theoretic models assume decision makers follow in responding to new information.[10] But why should this be true, since Kennedy could have issued a warning whether he was bluffing or not, and his decision to do so would likely depend on what effect he expected it to have on Khrushchev?

Probably the best indication of how far one can go in answering this question without a model is a brilliant book on this subject by Robert Jervis called *The Logic of Images in International Politics*. (Jervis 1970) Jervis distinguished between what he called "indices" and "signals." The difference between them is, roughly, that indices are indicators of a state's intentions or capabilities that cannot be faked, while "[b]oth the sender and perceiver realize that signals can be as easily issued by a deceiver as by an honest actor." (Jervis 1970, 21) Thus "...the question that arises is not, 'Why do states fail to believe others' signals in a crisis?' but 'Why do they ever believe signals in a crisis?' " (Jervis 1970, 96) Jervis provides examples of cases in which states did believe signals, but says, "The reasons for the surprising reliance on signals in these cases are not clear." (Jervis 1970, 102)

A model can provide a possible answer to this question. To construct such a model one represents choices confronted by individuals as branches on a tree, while the tree represents the sequence of choices as well as what

---

[9]Lebow's characterization of a rational choice is that it consists of processing "information in a relatively straightforward and honest manner in order to discover the best policy alternative." (Lebow 1981, 101)

[10]Another implication of Bayes' rule is that even after being warned, Khrushchev's estimate of how likely it is that Kennedy is bluffing would be influenced by how likely he thought it was before he was warned. Lebow attributes such reasoning to the " 'masking effect' of preexisting beliefs," and interprets it as evidence of faulty thinking. (Lebow 1981, 105)

each actor knows about the choices made by others when he or she chooses. The end points of branches to this tree are the outcomes the actors are interested in. Uncertain outcomes are represented by lotteries. It has been established that if actors have consistent preferences over all possible outcomes, including both certain and uncertain ones, then their choices will be equivalent to the choices that would be made by someone whose expectations can be represented by probabilities, whose relative evaluations of outcomes can be given numerical values (called utilities), and who make choices as though they were maximizing the product of these probabilities and utilities (or their "expected utility"). Thus subjective (or personal) probabilities and utilities are used to model expectations and evaluations in such a model. Since the outcome of one actor's choice depends on another actor's choice, these choices must be consistent both with each other, and with the probabilities that reflect the actors' expectations (and vice versa). Using such a model it is possible to show that as long as signals are not costless, actors might expect that someone who was bluffing would be less likely to issue a warning signal than someone who was not, which could explain why a signal could lead an adversary to revise his expectations of a state's future behavior.

Such a model of the Cuban Missile Crisis could thus provide an explanation of why, prior to putting those famous missiles into Cuba, Khrushchev believed an American military response to them was unlikely, but after he saw the US blockade of Cuba he revised his expectations and removed them. Jervis's distinction between signals and indices makes it hard to understand this, as the following passage reveals:

> Since the blockade revealed an American willingness to run some risks in opposing the Soviet venture it was partly an index.... But it also constituted a clear signal which could have been a bluff and did not involve any strong proof of American willingness to take the much higher risks which would have been necessary to secure American goals had Russia not retreated. (Jervis 1970, 22)

A model reveals how the risks entailed by the blockade gave credibility to a signal that it would otherwise not have had.[11]

However, a model of the Cuban Missile Crisis in particular might also reveal an entirely different possible explanation of what happened, one in which Khrushchev did not doubt Kennedy's willingness to use force but believed the use of force was all too likely, and hoped only to forestall it by putting missiles in Cuba before they were discovered. Thus what Khrushchev learned that led him to alter his behavior might have been not that Kennedy was more willing to use force than he had initially believed, but simply that the missiles had been discovered before they were completed, and therefore the game was up and he had better cut his losses while he could. This explanation differs from the first not primarily because it implies different objectives on Khrushchev's part, but because it implies different expectations about Kennedy's likely future behavior. A careful sifting of the evidence would be required to determine which is the better explanation of what happened, and inferences from each of the competing models can help one determine what evidence to look for.[12]

As this last example illustrates, one must distinguish among criticizing a particular model, criticizing the elements out of which an entire class of models might be constructed, and criticizing what Powell has called "the modeling enterprise." (Powell 1999) One can criticize a model of the Cuban Missile Crisis that implies that the American blockade was crucial in determining the outcome by constructing a model in which it was insignificant—the fact that one model is inadequate does not imply that an acceptable one could not be constructed. To argue that the inadequacies of a particular model imply that any use of models whatever is bound to be wrongheaded, however, would be like arguing that if one calculated one's travel time for a trip by using a map and got it wrong, one's mistake must have been in using a map.

But what about the elements from which the model is constructed? These are what constitute the "rationality" part of "rational choice theory." They include both elements taken from individual decision theory (subjective probabilities, utilities, and decisions that maximize expected utility) and elements taken from game theory (the concept of a perfect Bayesian equilibrium, and sufficient common knowledge to support it). These elements are, no doubt, often unrealistic. There are two things to bear in mind in assessing the significance of this fact. The first is, as already emphasized, that all models are inaccurate to some degree and in some respects—otherwise they would

---

[11]Walt used Jervis's book to buttress his claim that formal models have not really added much to our understanding of international politics, claiming that the concept of "costly signals" just outlined is "virtually identical to Robert Jervis's distinction between 'signals' and 'indices,' which he laid out more than twenty-five years ago." (Walt 1999, 29) This is simply untrue, as the example just quoted illustrates. The blockade of Cuba was not, as Walt quotes Jervis as saying an index would have to be, "behavior that is felt to be too important or costly in its own right to be used for other ends," since it could have been used as a signal as well as a way of preventing further shipments to Cuba. And Jervis leaves entirely unclear how any risks associated with the block-

ade could enhance its credibility as a signal. (Michael Spence, who pioneered the formal analysis of costly signaling, actually took Jervis's work as a point of departure—see Spence 1974, 10–11.)

[12]Thus the notion that "rational choice theory" is intended to be a "universal deductive theory," as Shapiro claims, is utterly without foundation. (Shapiro 2000) Moreover, there is no incompatibility between constructing models of this sort and doing historical research; indeed, they complement each other.

not be models. The question is whether a model that is more accurate in some respect or other would lead to a different inference. Sometimes the only way to determine the answer to this question is to construct what might be a better model to see if that makes a difference. This is done all the time, and does not imply that we would be better off if we abandoned the use of models entirely.[13]

Second, what critics of "rational choice theory" call "rationality" is in fact a whole set of assumptions about the way people make choices, many of which could be altered quite independently of others. This is why people who actually construct "rational choice" models do not usually talk very much about rationality, but instead just construct models employing some specific set of assumptions. Moreover, eliminating the assumptions that fall under the heading of "rationality" without replacing them with something else would make valid inferences impossible. Thus anyone who wants to explain human choices but is dissatisfied with the set of assumptions commonly lumped under the heading of "rationality" must find an alternative set of assumptions to fill the gap that would be left if they were eliminated.[14]

As I pointed out earlier, the basic underlying assumption employed in models of individual rationality is that individuals have consistent preferences. The justification for associating this with the term "rationality" is that arguments can be given to the effect that if people thought carefully about what they were doing they would want their preferences to be consistent in the ways that are assumed by expected utility theory, and would change their behavior if an inconsistency were pointed out to them. Thus these assumptions represent what might be called, to borrow a term used by John Rawls, a "reflective equilibrium." Many deviations from such an equilibrium might be random, and thus corrections for them might have to be tailored for individual cases (as your guests will have to adapt to kinks in the roadway that you neglected to include on your map). To do more than this one requires some reason to believe that certain kinds of deviations from equilibrium are common in certain kinds of circumstances.[15]

The rationale for the elements of these models derived from game theory is less clear, as is their relation to the elements derived from decision theory, and therefore they are more controversial. Nash's concept of an equilibrium

has already been modified in a variety of ways, beginning with Selten's notion of a "perfect equilibrium," and Harsanyi showed how the amount of common knowledge assumed by game theory as originally formulated could be drastically reduced. (For these contributions Harsanyi and Selten shared the Nobel prize in economics with John Nash.) Nonetheless, this remains an active area of research and controversy, and therefore it would be a mistake to assume that "rational choice theory" necessarily entails a commitment to any very specific set of assumptions about these matters other than a mutual recognition of the interdependence of choices.[16]

I hope by now that it is clear that the common belief that "rational choice theory" is a theory of human behavior that predicts what people will do in various situations is simply a misconception. It is not "rational choice theory" that explains people's choices, but models constructed using the elements just described that do so, and if a model fails to explain some choices of interest it is the model that has failed and not "rational choice theory" (unless, of course, no model constructed from any of these elements could possibly explain the choices in question).[17] The proper response to the failure of a model is to try to construct a more successful one, in which case the failed model has done its job by showing that such an explanation will not work. Thus models that are wrong can nonetheless be a contribution to knowledge, and one should be pleased to have constructed an explanation that can actually be shown to be wrong. Otherwise one's work is, to use words commonly attributed to the physicist Wolfgang Pauli, "not even wrong."

Consider, for example, the problem of explaining why people vote. Shapiro says that "egoistic rational maximizers have no reason to vote, given the infinitesimal odds of affecting the result," and therefore, he claims, the fact that people do vote is inconsistent with "rational choice theory." "Rational choice theorists," he says, respond to this by introducing ad hoc complications to their theory rather than accepting the fact that the theory is wrong. (Shapiro 2000) But "rational choice theory" does not say that egoistic rational maximizers would vote only if there is a non-negligible chance that they will be able to influence the outcome of an election, that is merely what some models assume.[18] And the notion that the reason people vote is that they want to have an influence on the outcome

---

[13]See Powell's discussion of "modeling dialogs" in Powell 1999, especially chapters one and six.

[14]In addition, as the Cuban Missile Crisis example illustrates, it is not possible to evaluate the effect of departures from rationality without knowing what rational individuals would do.

[15]The importance of prospect theory is that it claims to have identified such patterns in the ways in which people respond to uncertainty. (However, contrary to what many people believe, not everything in prospect theory is inconsistent with the axioms of expected utility theory.) A recent survey of attempts to find regularities of this sort can be found in Starmer 2000. For a discussion of the implications of this work for modeling choices, see Roth 1996.

[16]For an interesting discussion of recent contributions in this area and their possible relevance to understanding international politics, see O'Neill 1999.

[17]Thus the term "rational choice theory" is itself the source of much confusion and therefore best avoided, which is why I have used it only in quotation marks.

[18]Note that words like "egoistic" and "self-interested" do not have clearly defined meanings. "Rational choice" models assume that individuals have consistent preferences. The preferences assumed in any particular model might or might not deserve to be called egoistic or self-interested.

of an election is not confined to "rational choice theory." By making clear how difficult it is to explain voting in that way, such models are a stimulus to constructing more adequate explanations of voting. This is not, as Shapiro says, a way of "saving the models at the price of rendering them banal," but a way of advancing our understanding of why people vote—something that cannot be done merely by determining the correlates of voting.

Just do it

For reasons that are not clear to me, political scientists spend about as much time talking about talking about politics as they do talking about politics. Some of this talk is influenced by second and third hand accounts of debates in the philosophy of science, as though figuring out how to study politics were the job of philosophers. One wonders how physicists managed to get anything done, and Darwin was able to invent the theory of evolution, before philosophers of science were available to tell them what to do.

This preoccupation with the philosophy of science has things exactly backwards: the problem faced by philosophers of science is to give a coherent account of what people do when they explain things, not to tell them how to do it. Our problem is to do it, and we do not need philosophers of science to tell us how. We merely need to be sure that the conclusions we reach follow from the premises offered to support them, to be imaginative in thinking of premises from which facts that we are interested in might be derived, to be vigilant in ruling out alternative explanations, and to be willing to do whatever is necessary to achieve those goals. Often, but not always, mathematical models are indispensable tools for achieving them. Some, but not all, of the techniques for creating these indispensable mathematical models constitute what is commonly referred to as "rational choice theory."

One source of confusion about all this has been what was called the behavioral revolution, whose victors had the word "science" prominently displayed on their banners, and tended to confuse the general meaning of "explanation" with the way the word is used in statistics.[19] This has led even people who would not know what to do with a variable if one were thrown in their laps to speak of independent and dependent variables when offering an explanation of something, and to confuse attempts to explain individual events such as the Cuban Missile Crisis with "small N studies."[20] It has also led to endless debates about whether political science should be, or can be, a science.[21]

Inept debates about how to talk about politics lead to seemingly irreconcilable conflicts among political scientists, which make it impossible for departments to agree on what graduate students should study. This places an enormous burden on the graduate students, who are left with the impression that before they can begin they must find their own answers to questions that professional political scientists are unable to agree on the answers to. Some conclude that political science consists largely of debates about the relative merits of competing "paradigms." And just as politicians in the US may "play the race card" to create conflicts from which they can derive political advantage, or Slobodan Milosevic "played the ethnic card" in Yugoslavia for the same reason, there are political scientists who "play the rational choice card" to create departmental conflicts from which they hope to benefit.

Often "rational choice theorists" are accused of "intellectual imperialism," and tolerance is said to be the remedy for these conflicts. The clear implication of this recommendation is that graduate students, in deciding what to study, should just pick any "approach" or "paradigm" they find congenial, since they are all equally deserving, and we should let a hundred flowers bloom. This reasoning implies that for most students it should be a "no brainer" to avoid "rational choice theory" like the plague, since it is both difficult and controversial. Graduate study already appears hard—why add to one's burdens unnecessarily?

This is the wrong response to conflict. The one thing we should not tolerate is sloppy thinking, and we should be prepared to do whatever is necessary to avoid it—if this is intellectual imperialism, then so be it. You should check the logical validity of arguments in the literature, and strive to attain it in your own. Mathematical models are not always necessary for this purpose, but often they are, and when they are, everyone should be prepared to use them. You should therefore learn how to construct them, and how to evaluate them. And you cannot rely on the nursery tales found in second and third hand accounts of something called "rational choice theory" to tell you how.

---

know what their dependent variable is." This statement baffled me for a long time. I'm afraid I still don't know what my dependent variable is, nor do I remember what he said his was.

[21]Is paleontology a science? Evolutionary biology? Evolutionary psychology? The question is not well defined, and the answer is not to be found in a statistics book. Nor is it very interesting. The important question is whether there is any reason to believe what people who work in those areas tell us.

---

[19]The "behavioral revolution" was just the use of statistical methods to analyze quantitative data about political behavior. It may be hard for some to believe that this was once very controversial.

[20]When I first started teaching, a colleague who was a product of an avowedly "behavioral" political science department said to me that "the problem with people who study international politics is that they don't

REFERENCES

Allison, Graham T. (1971) *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown and Company.

Bates, Robert H. (1997) "Comparative Politics and Rational Choice: A Review Essay." *American Political Science Review* 91:699–704.

Bensel, Richard Franklin. (1990) *Yankee Leviathan: The Origins of Central State Authority in America, 1859–1877*. New York: Cambridge University Press.

Green, Donald P., and Shapiro, Ian. (1994) *Pathologies of Rational Choice Theory*. New Haven: Yale University Press.

Harman, Gilbert H. (1965) "The Inference to the Best Explanation." *The Philosophical Review.* 74:88–95.

Howson, Colin, and Urbach, Peter. (1993) *Scientific Reasoning: The Bayesian Approach*, 2d. ed. Chicago: Open Court Publishing Company.

Jervis, Robert. (1970) *The Logic of Images in International Relations.* Princeton: Princeton University Press.

Lave, Charles A., and March, James G. (1993) *An Introduction to Models in the Social Sciences*. Lanham, Maryland: University Press of America.

Lebow, Richard Ned. (1981) *Between Peace and War: The Nature of International Crisis*. Baltimore: The Johns Hopkins University Press.

Lipton, Peter. (1991) *Inference to the Best Explanation.* London and New York: Routledge.

Mayer, Ernst. (1991) *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought.* Cambridge, Massachusetts: Harvard University Press.

Munck, Gerardo L. (2001) "Game Theory and Comparative Politics: New Perspectives and Old Concerns." *World Politics* 53:173–204.

O'Neill, Barry. (1999) *Honor, Symbols, and War*. Ann Arbor: The University of Michigan Press.

Powell, Robert. (1999) *In the Shadow of Power: States and Strategies in International Politics*. Princeton: Princeton University Press.

Roth, A. E. (1996) "Comment." In Kenneth J. Arrow, et al., eds., *The Rational Foundations of Economic Behavior.* New York: St. Martin's Press, Inc.

Shapiro, Ian. (2000) "A Model That Pretends to Explain Everything," *The New York Times*, February 26, 2000.

Spence, A. M. (1974) *Market Signaling: Informational Transfer in Hiring and Related Screening Processes.* Cambridge, MA: Harvard University Press.

Starmer, C. (2000) "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal of Economic Literature* 38:332–382.

Strawson, P. F. (1952) *Introduction to Logical Theory*. London: Methuen & Co. Ltd.

Wagner, R. Harrison. (2001) "Bargaining, War, and Alliances." Paper presented at the Annual Meetings of the American Political Science Association, San Francisco, CA, September, 2001.

Walt, S. M. (1999) "Rigor or Mortis? Rational Choice and Security Studies," *International Security* 23:5–48.

Waltz, Kenneth. (1979) *Theory of International Politics*. Reading, Massachusetts: Addison Wesley Publishing Company.