*Jeffrey Friedman*

# *SYSTEM EFFECTS* AND THE PROBLEM OF PREDICTION

ABSTRACT: *Robert Jervis's* System Effects *(1997) shares a great deal with game theory, complex-systems theory, and systems theory in international relations, yet it transcends them all by taking account of the role of ideas in human behavior. The ideational element inserts unpredictability into Jervis's understanding of system effects. Each member of a ''system'' of interrelated actors interprets her situation to require certain actions based on the effects these will cause among other members of the system, but these other actors' responses to one's action will be based on their own perceptions of their situation and their interpretations of what it requires. These ideas are fallible, but we cannot predict the mistakes people will make if the errors are based on information we do not have or do not interpret in the same way they do. Not only members of a system but social-scientific observers and policy makers are ignorant of others' information and interpretations, and therefore are as likely to err in their behavioral predictions as are members of the system. Thus, Jervis's book raises serious questions about how to evaluate policies directed toward producing positive system effects. The questions are unanswerable at this point, but they might be susceptible to analysis by an ambitious form of political theory.*

Robert Jervis's *System Effects: Complexity in Political and Social Life* (Princeton University Press, 1997) is a classic work of political science that, unlike several that have previously been the subject of symposia here—Philip E. Converse's ''The Nature of Belief Systems in Mass

Jeffrey Friedman, edcritrev@gmail.com, Critical Review Foundation, P.O. Box 869, Helotes, TX 78023, thanks Samuel DeCanio, Stephen DeCanio, and Nuno Monteiro for comments on previous drafts.

*Critical Review* 24(3): 291–312 ISSN 0891-3811 print, 1933-8007 online
 http://dx.doi.org/10.1080/08913811.2012.779806

Publics'' (1964), Jeffrey K. Tulis's *The Rhetorical Presidency* (1987), and Philip E. Tetlock's *Expert Political Judgment* (2005)[1]—did not come close to having the impact it should have had.

The problem may have been that the book was a poor fit with the subdisciplinary organization of political science. Jervis was a renowned scholar of international relations when he published this book; it was explicitly presented as an outgrowth of the structural neorealist international-relations theory of Kenneth Waltz (Jervis 1997, ch. 3); and most of Jervis's examples of system effects are historical incidents and developments in international affairs—that is, examples of the interaction of states, conceived as parts of an international system of states. However, as Nuno Monteiro (2012) points out below, international-systems theories do not satisfy the evidentiary criteria that international-relations scholars had begun to adopt when the book was published.

Moreover, while the book is unmistakably that of a scholar of international relations, it freely crosses sub-disciplinary (and disciplinary) boundaries, drawing from public policy, American politics, and domestic politics to develop the key themes of systemic nonlinearities, feedback effects, indirect effects, and the effects of contingency, all of which are building blocks of complex-systems theory. It is only natural, then, to read the book as an application to politics of complex-systems theory (as opposed to international-systems theory), as Andrea Jones-Rooy and Scott E. Page (2012) do below.[2]

## *Jervis as Complex-Systems Theorist*

However, Jervis builds something rather unusual from the standard elements of complex-systems theory. For example, he writes that in politics, ''differences in expectations and policy preferences are often rooted in different beliefs about feedbacks'' (Jervis 1997, 130), and he spends much of his chapter 4 discussing how, in turn, these different beliefs affect political behavior through their influence on preferences and expectations. And among the phenomena caught in Jervis's net are interaction effects, which are not so readily fit into the complex-systems framework, since, as Jervis defines them, they change ''the environment of action, so that other actors do not respond as anticipated'' (Jervis 2012, 395). When anticipations are frustrated, as when expectations and preferences are formed (as opposed to being posited by the theorist),

we cannot be discussing the interactions of thoughtless biological organisms or mindlessly rule-following automata (Mitchell 2009), which are the actors in classic systems-theory computer modeling (Page 2011, 42–43). ''System effects can occur with inanimate objects,'' Jervis (1997, 253) observes, ''but greater complexities are introduced with human beings whose behavior is influenced by their expectations of what others will do, who realize that others are influenced by their expectations of the actor's likely behavior, and who have their own ideas about system effects.''

This passage, and the many strategic interactions discussed by Jervis, convey the game-theoretic dimension of the book, in which systemically connected adversaries try to outguess each other. Game theory has as much affinity as complex-systems theory with Jervis's project. Yet game theory still misses a crucial element, because it requires the modeler to predict the players' reactions to each other's actions. Jervis, however, contends that ''to claim that we can be certain of how each actor will respond, how the different behaviors will interact, and how people will then adjust to the changed circumstances goes beyond the knowledge we can have'' (Jervis 1997, 72).

The key to understanding this contention is, I think, found in Jervis's previously quoted claim that human beings introduce extra complexities because they ''have their own ideas about system effects.'' Arguably, this is because people have access to different information about whatever systems they are entangled in, and because they interpret this information differently. ''Few social acts fail to alter the informational as well as the physical environment,'' Jervis (1997, 145) writes, and no given datum is known to all or understood in the same terms by all. Thus, ''disagreements [about the likely consequences of an action] are not surprising'' (ibid., 73), and the divergent information streams and interpretive responses that lead different agents to disagree about the consequences of their actions—i.e., to disagree about their actions' likely system effects—cannot be transcended by those who are trying to predict the agents' behavior. The observers at Time 1, like the agents themselves, are trying to discern the best action for the agents to take, and in both cases this depends on forecasts of various possible actions' results—in the form of other actors' reactions—at Time 2. But since neither actors nor observers can know in advance how other actors in the system will perceive, interpret, and (after again forecasting the results of various actions)

respond to $T_1$ actions, the reliability of actors'/observers' forecasts about behavior at $T_2$ is questionable at best.

## *Jervis as Political Scientist*

Accordingly, the book is notably retrospective, as Richard A. Posner (2012) emphasizes below. Jervis issues no predictions and constructs no formal models.[3] He does adduce plentiful examples of various patterns of interactive behavior and mistake, but he is not constructing a political "science" in the sense that he assumes that he is describing universal behavioral laws.

In perhaps the most formalistic and "law-like" passage in the book (one that is of no great importance to the overall argument), Jervis (1997, 216) writes that states are

> likely to develop good relations with each other if they share allies. It will be difficult for both *A* and *C* to maintain their close ties with *B* if they are adversaries. If the alliance with *B* is highly valued, *A* and *C* will have to mute their quarrels, and the desire to increase the coalition's strength will give both *A* and *C* incentives to bring the other into the fold. There will be a cost, however, if the main target of the alliance between *A* and *B* is not the same as that between *C* and *B*. In this case, if *A* and *C* join together they will not only bolster each other, which they will resist doing if they have bilateral conflicts, but also will pay the price of taking on the other's adversaries. For these reasons, the impact of having a common friend is less than that of having a common adversary; systems are more likely to be unbalanced by not having those with common allies be themselves allies than they are by having states fail to unite in the face of a common adversary.

However, even when Jervis draws generalizations like this one, they are expressed as mere "likelihoods," and they are clearly conditioned upon motives and knowledge that may not be present in a given case, and that may be counteracted by other factors if they are present. In short, they are Weberian ideal types: logically coherent explanations of what will happen *if* (1) certain presuppositions hold, and (2) other factors do not intervene.

While it is typical for social scientists to attach *ceteris paribus* clauses to their predictions, meeting Weber's second criterion, Jervis takes the clause seriously enough to notice that it rarely holds good, rendering prediction unreliable. Perhaps more importantly, where most social scientists ignore

the first criterion, Jervis does not. That is, he does not assume that the preconditions of an ideal type will be applicable to all cases *as long as* the *ceteris paribus* clause holds good.

The assumption of uniformity in underlying causes allows incautious social scientists to infer future behavior from past behavior (as long as the *ceteris paribus* clause holds good). But Weber ([1904] 1949) viewed the purpose of social science as purely historical, and the purpose of historical research as the discovery of *whether* an ideal type's preconditions are applicable to a given case. If, by contrast, ideal types expressed lawlike regularities, then historical research could only be directed toward seeing if, in a given case, the underlying tendency was overridden by some other factor (violating the *ceteris paribus* clause). Weber's own historical research culminated in a bold, bleak understanding of modernity as the hyper-rationalized product of contingent chains of events in intellectual and cultural history (especially the history of religion). Since there was no underlying necessity to these events, they could not be used as the basis for predictions of the future. But in that case, what was the point of social science, i.e., historical research? To clarify the nature and roots of our present condition.

A more typical Jervisian statement immediately follows the passage I have extracted above:

> States can also manipulate the dynamics that produce consistency in order to produce a desired alignment: If a state wants to make or solidify an alliance with another, it may pick a quarrel with the other's adversary. In the previous chapter, I noted A. J. P. Taylor's argument that in order to draw France into his orbit, Bismarck created a colonial conflict with France's adversary, Great Britain. Greater evidence supports the view that one reason why Anwar Sadat distanced himself from the USSR in late 1972 was his belief that this would bring financial assistance from the conservative and anticommunist Arab oil kingdoms. While Nixon and Kissinger did not create the crisis between India and Pakistan in 1971, they did realize that the resulting friction between the U.S. and India would increase the common interest between the U.S. and China. The same reasoning explains why British leaders foresaw some gain in the Ottoman Empire's siding with Germany in 1914: "The best method of persuading the Balkan States to join the Allies would be alliance against their common and traditional enemy, the Turk." (Jervis 1997, 216)

What is being expressed in these examples are fruitions of an underlying *possibility*, not the operation of an underlying law. The aim of Jervis's

retrospective focus, however, is not entirely Weberian. Although Jervis does use ideal types to explain the past, he also goes a step further by pointing out the applicability of given ideal types to *many* actual historical situations. He is looking for regularities, not trying to explain (except incidentally) how we got to where we are.

Still, in his treatment of the regularities he finds, he does not take the fatal additional step of claiming that they are products of knowable universal laws that would enable the prediction of future cases (except inasmuch as we can always predict that *if* the presuppositions of an ideal type *do* apply in a future case and are not counteracted, *then* we can expect certain results).

It is true that by backing away from the enunciation of universal laws, Jervis remains in line with most political scientists' *explicit* agendas; only a minority of empirical political scientists claim that their findings can be duplicated everywhere and always, such that they can be used to predict the future.[4] International relations is the most predictively oriented subfield of political science (Monteiro and Ruby 2009); here Jervis's book makes a striking contrast. Yet even in the other empirical subfields, the positivist notion that everything must ultimately be reducible to (knowable) universal laws displays its hold in excrescences such as quadrennial attempts to derive formulae for predicting the next presidential election outcome, usually on the basis of "real" (economic) factors.[5] Even if one follows Milton Friedman (1953) in insisting that the factors expressed by such formulae are not supposed to be *actually* causing electoral outcomes, but are merely variables that (for some unknown reason) allow us to make good behavioral predictions, in practice one usually wants to know what *is* actually causing the behavior, and it is all too easy to assume that whatever is causing it—since it seems to be responsible for a behavioral regularity—must be some universal human disposition.

Indeed, one would only think to test factors such as economic growth as predictors of voting behavior because one finds it plausible that voters are, at bottom—past, present, and future—members of the species *Homo economicus*. However, when one moves from treating objective economic variables as mere predictive instruments to treating them as "real" causal factors, one is tacitly assuming that voters are somehow able to *know* the economic variables; if they did not, for example, know that there was a recession, they could not respond to it by voting against the incumbent.[6] (This applies even to retrospective-voting models that are intended to

minimize voters' need for information.)[7] But aggregate economic factors are not self-evident to anyone; they must be mediated. How would even an unemployed voter know whether the business she used to work for had closed because of a recession or, alternatively, because of management errors or some combination of factors? The very concept of a recession—like the questions of whether a recession is occurring and why—must somehow be theorized and communicated to voters; one cannot directly observe a recession, let alone identify its sources, but instead must hypothesize that certain accessible facts (such as one's own unemployment, or an unemployment figure heard on the news) are politically relevant because of government actions or inaction that caused or allowed the "recession." The causal force behind regularities such as those identified by election forecasters must therefore be voters' subjective ideas, however inchoate, about objective economic conditions and their causes—not the conditions and causes themselves.

To the extent that subjective ideas are mistaken, models based on "real" factors will produce equally mistaken forecasts—i.e., they will be unrealistic.[8] Thomas Holbrook and James C. Garand (1996) suggest that most voters in the 1992 election severely overestimated the straits the economy was in. Marc J. Hetherington (1996) shows that during the 1992 campaign, the media failed adequately to report that the recession that had occurred under the incumbent had ended twenty months before the election, and that voters who were most heavily exposed to the media had the most inaccurate impressions about economic performance. So the incumbent lost, and the opposition candidate's slogan was "It's the Economy, Stupid." Such empirical anomalies (from the perspective of those who view objective factors as determinative) get washed out by the aggregates used by the forecasters.[9]

The assumption that any regularity must reflect universal tendencies may also lead to *illogical* answers to the question of what those tendencies really represent. An example is the fact that rational-ignorance theory is often casually thought to explain twentieth-century U.S. voters' political ignorance. According to this theory, the regularly observed fact that most twentieth-century American voters show little awareness of political events, personages, and policy debates can be explained by the fact that voters realize that any one vote is unlikely to affect the outcome of an election with many voters, so they rationally calculate that it would be a waste of time for them to pay attention to political news. However, if they *knew* that their votes were unlikely to affect the outcome, and if

their motive in voting were to affect the outcome, then they would abstain from voting rather than deliberately underinform themselves before voting.[10]

The massive reality of voting thus precludes the applicability of rational-ignorance theory to anyone but *nonvoters*. Yet the theory persists. It is hard to avoid the conclusion that it persists because of an unexamined slide from the identification of patterns of political ignorance in recent American political history to the conclusion that since all political behavior is lawlike, any regularity implies a knowable, universal cause.[11] Such a cause must, by definition, be an objective fact—whether the unemployment rate or the odds against one's vote mattering in a large electorate. Yet objective facts can influence one's conscious decisions only if one subjectively perceives them and interprets them as germane to one's decision.

Thus, all positivism will in one way or another have to minimize the importance of subjective beliefs. The usual means to this end is to assume tacitly that the ideas of the agents whose behavior one is trying to predict (or retrodict) are identical to one's own.[12] Thus, electoral predictions based on economic statistics, and explanations of voter ignorance by means of rational choice, both minimize subjectivity by assuming, in effect, that voters know the facts the theorist takes to be crucial (economic variables or the odds that a vote will matter); and by further assuming that the voters know, and agree with, the theories that make the political scientists think that those facts "should" in some sense decide the agents' actions. Yet positivists routinely fail to identify mechanisms by which these facts and interpretations might plausibly be *communicated* to the agents whose actions are being predicted—and communicated in precisely the same form that has made them persuasive to the positivist scholar. Technically speaking, positivism—by which I simply mean predictive social science—substitutes ontology (objective facts) for epistemology (some mechanism by which agents can perceive objective facts and interpret their significance for agents' future actions).

## *Jervis as Political Epistemologist*

In *Perceptions and Misperceptions in International Politics* (1976), Jervis pointed out that in the study of international relations, too, scholars routinely "explain and predict" actions by inferring the appropriate

behavior of an agent confronting given objective environmental factors (ibid., 16). In response, Jervis observed that foreign-policy decision makers could themselves could use the same method of inference: If they

> believed that the [objective] setting is crucial they would not need to scrutinize the details of [another] state's recent behavior or try to understand the goals and beliefs held by the state's decision-makers. It would be fruitless and pointless to ask what the state's intentions are if its behavior is determined by the situation in which it finds itself. Instead, observers would try to predict how the context will change because this will tell them what the state's response will be. Decision-makers could then freely employ their powers of vicarious identification and simply ask themselves how they would act if they were in the other's shoes. They would not have to worry about the possibility that the other might have values and beliefs that differed from theirs. It is interesting, although not decisive, to note that decision-makers rarely feel confident about using this method. They usually believe both that others may not believe as they would and that the decision-makers within the other state differ among themselves. So they generally seek a great deal of information about the views of each significant person in the other country. (Ibid.)

Presumably Jervis allows that this type of information seeking by decision makers does not decisively refute the notion that behavior can be predicted from objective circumstances because the decision makers could be wrong: Contrary to their belief that subjective perceptions cannot be read unambiguously from objective circumstances, it could be that these circumstances are so obvious to all that "decision-makers usually perceive the world quite accurately" (Jervis 1976, 3), as the social scientists assume. However, if the decision makers of country A are trying to infer country B's decision makers' subjective perceptions of objective conditions, it can only be because the decision makers in country A think that the perceptions of those in country B may differ from objective reality (as interpreted by those in country A). From the social scientists' perspective, this constitutes a serious mistake by the actors in country A whose behavior they are trying to predict. It is a mistake not only about the importance of the subjective perceptions of country B's decision makers, but (therefore) about the *objective* environment A's decision makers face, which consists, in large part, of the behavior and potential behavior of other states' decision makers. It will not do, then, to predict the behavior of A from the objective environment, since A's decision makers evidently fail to understand

that environment—falsifying the assumption that the environment is unambiguous to all.

Because *System Effects* is almost entirely retrospective, Jervis is able to avoid the logical fallacies and empirical overgeneralizations that so many of his colleagues commit. But one might also say that the book *had* to be retrospective because of the crucial element that Jervis adds to game theory and complex-systems theory: attention to the unpredictable subjective beliefs about objective reality that are prerequisites for political action. To the extent that beliefs govern behavior, and to the extent that they are unpredictable, then clearly we cannot predict future behavior. We may be able to infer past beliefs from past behavior by applying an ideal typology to the behavior. But we will be unable to predict the applicability to a future situation of an ideal type that is contingent on actors' beliefs. Thus, reliable forecasts of human behavior would appear to be out of bounds.

Why, however, should we think that subjective beliefs are unpredictable? Building on Jervis's 1976 argument, one might first of all note that reasonable people could not *disagree* about the best course of action in a given circumstance if that course of action could be unambiguously derived from the facts alone. We have already seen that Jervis has shown that foreign-policy decision makers disagree with international-relations scholars who claim that their adversaries' actions can be unambiguously derived from the facts alone. If scholars and policy makers can disagree, cannot policy makers themselves disagree?

Much of *System Effects* is devoted to exemplifying just such disagreement. For instance, Jervis (1997, 45) notes that "many cases of intelligence failure are mutual—i.e., they are failures by the side that took the initiative as well as by the state that was taken by surprise." We can stop right there: A state could hardly be surprised by another state's actions if the decision makers and intelligence analysts of the two states agreed with each other on the objective facts and on how to interpret them. Yet "the U.S. did not expect the Russians to put missiles into Cuba or Japan to attack Pearl Harbor because American officials knew that the U.S. would thwart these measures if they were taken. These judgments were correct, but because the other countries saw the world and the U.S. less accurately, the American predictions were also inaccurate" (ibid.).

Decision makers may also disagree among themselves, not just with their counterparts in other states. For example,

> [an] agent's beliefs about what tactics are appropriate may . . . differ from the views of those at home. The man on the spot almost always feels he knows more about the local situation than his superior and believes many of his instructions to be hopelessly out of touch with the reality he sees. His superiors, he is apt to conclude, do not understand what is happening or what can be achieved. . . .
>
> An agent's disobedience can take various forms. In some cases an agent may refuse to deliver a message or may substitute one of his own for that of his government. In negotiations with Portugal in 1943 George Kennan gave the Portuguese government an assurance that was "in direct violation of the written orders I had in my safe." . . . In 1809 the British minister to the United States broke his instructions and signed a treaty with America that did not meet major British demands. (Jervis 1976, 332–33)

Jervis also adduces a case that might have had a significant effect on Hitler's expansionism: the British ambassador to Germany appended to his official protest against the *Anschluss* the caveat that Austria "'had acted with precipitate folly.'" This may not only have encouraged Hitler, if he misunderstood it as a signal of the British government's real attitude; it may have discouraged the British government from taking further measures to express its resolve against Nazi expansionism, since the ambassador did not report his remarks to his superiors (ibid., 333–34).

Given that one's best course of action in a system will depend on its consequences once other actors perceive one's action and respond to it (if they do); and assuming, for the sake of argument, shared values among the decision makers;[13] then all decision makers would derive the best course of action from objective circumstances if this could be done unambiguously. In that case, there would be no disagreement among decision makers. Thus, unless one minimizes the possibility that the actual disagreement that we routinely encounter is *reasonable*—e.g., by ascribing disagreement to irrational emotions, which could perhaps be forecast (with the assistance of psychology)—the project of predicting human behavior faces a serious obstacle in what might be called "the fact of genuine disagreement." Genuine disagreement among reasonable decision makers who are aiming at the same end is as anomalous for those who would predict human behavior as voting is for those who would attribute voters' political ignorance to their knowledge that their individual vote does not matter.

Disagreement about the best course of action also indicates that one or both of the parties to the disagreement is in error, as is clearly evident in

Jervis's take on Soviet and American predictions of each other's actions before the Cuban Missile Crisis and Japanese and American predictions of each other's actions before Pearl Harbor. If we could predict people's beliefs and thus their actions from their objective circumstances, such errors would be impossible. A world of predictable human beings would be a world of infallible human beings. Our predictions will be frustrated to the extent that mistakes are made by those whose actions we would presume to predict—unless we share in the same mistaken logic or faulty information stream that might have led those particular actors to err. But if the content of the information to which one is exposed varies from person to person or changes over time, we cannot possibly share in the information streams that will be available to those whose actions we try to predict. To the extent that their subjective perceptions are based on those streams, we will not be able to predict their perceptions, hence their actions.

Finally there is a fact that receives great emphasis in *System Effects*. Those whose behavior we might want to predict are not just trying to outguess each other's strategic moves; they are, more generally, trying to forecast what will happen when their actions, and others' actions, react against, and contribute to, the objective realities that everyone in a system is trying to perceive and predict. In short, the objective realities themselves change as actors interpret them and take actions based on their interpretations. A belief-dependent objective reality can no more be forecast than can the beliefs themselves.

This raises the question of what human "systems" really are. They can be loosely defined as webs of interacting agents, but the agents' interactions derive from perceptions of reality, not from some other (ontological) dimension of reality. If a "state" exists deep in the Amazon, unknown to the outside world, it is ontologically real but it is not part of "the state system." If a state that is part of that system (because it is perceived as such) builds a nuclear reactor for military purposes but portrays it as a peaceful source of electric power, there is an "interaction" only if analysts in other states see through the ruse, or suspect that it might be a ruse. Yet their suspicions may be unfounded—the perception of a threat may be mistaken—in which case the interaction is grounded not in ontology but in an erroneous theory held by the decision makers who misperceive the reactor as a threat. This does not, of course, make the interaction itself unreal. On the contrary: perceptions are real, and they can lead to real actions.

If one were to diagram these interactions, as Jones-Rooy and Page diagram various networks, all the action would be in the arrows between agents, but these would express *the observer's* theories about the theories agents have, or should have, about each other's theories (as well as "real" factors such as their military capabilities), and therefore about each other's likely actions. While subjective perceptions are objective realities in themselves, and while the objects of these perceptions may have "real" correlates, including actions, the system of interactions is an expression only of epistemological relationships, not extra-ideational realities.[14]

The "complexity" of the human environment to which Jervis refers may therefore best be seen as a product of the limits on our ability to understand and especially predict it, not as a product of its inherent qualities, such as nonlinearities, feedback effects, indirect effects, and the effects of contingency.[15] This is to suggest further that what makes a human system complex is not its nonlinear emergence as such, nor its "spontaneity," nor the number of factors that interact within it.[16] It is the difficulty of understanding and predicting the ideas of other people—those with whom we directly and indirectly interact, and those whom we, as scholars, observe or imagine directly and indirectly interacting.

## *Jervis as Political Theorist*

System effects, in this view, are caused by the fallible perceptions of people who are each trying to formulate accurate predictions of others' behavior based on their interpretations of what seems to them relevant evidence about each other's beliefs. Disagreements among agents who are pursuing the same goal indicate error on the part of some or all of them (although, given the limits of human knowledge, consensus would not necessarily indicate the absence of error). In turn, the fact that Jervis builds human error into his understanding of society puts him at odds with those who would predict human behavior: not just social scientists, but policy makers; and not just foreign-policy makers, but domestic-policy makers—be they voters, elected officials, appointed officials, or bureaucrats. They, too, are (sometimes unwittingly) predicting the effect of proposed actions on a human system (a polity). Thus, while Jervis's brand of political epistemology makes an uneasy fit with international-relations

theory, complex-systems theory, and game theory, it may yet provide instruction to normative political theorists.

Jervis (1997, 60) summarizes the upshot of fallible policy predictions (and of simplistic attempts to understand the roots of policy failures after the fact) by saying that "intentions and outcomes often are very different, regulation is prone to misfire, and our standard methodologies are not likely to capture the dynamics at work." Indeed, *System Effects* opens with an epigraph from a *New Yorker* story about marine biologist Sylvia A. Earle, in which Jervis (ibid., 3) ironically juxtaposes Earle's conviction that one needs to be aware of "the continuing interconnectedness of the system" (White 1989, 56) against her declaration that, to minimize oil spills, we should "mandate double-hulled vessels and compartments in tankers" (ibid., 46). Jervis points out that

> it seems obvious that if tankers had double hulls, there would be fewer oil spills. But interconnections mean that the obvious and immediate effect might not be the dominant one. The straightforward argument compares two worlds, one with single-hulled tankers and one with double-hulled ones, holding everything else constant. But in a system, *everything else will not remain constant*. The shipping companies, forced to purchase more expensive tankers, might cut expenditures on other safety measures, in part because of the greater protection supplied by the double hulls. The relative cost of alternative means of transporting oil would decrease, perhaps moving spills from the seas to the areas traversed by the new pipelines. But even tanker spills might not decrease. The current trade-off between costs and spills may reflect the preferences of shippers and captains, who might take advantage of the greater safety by going faster and taking more chances. If double hulls led to even a slight increase in the price of oil, many other consequences could follow, from greater conservation, to increased uses of alternative fuels, to hardship for the poor. (Ibid., 8)

Notice that the barrier to accurate prediction here is purely epistemological. If we could—in advance—read the minds of the agents about whom Jervis is speculating, there would be no problem in deciding whether to mandate double hulls.

Jervis is using standard microeconomic analysis of cost-benefit decision making—yet the analysis is couched as speculation. An economist could assert that, *ceteris paribus*, the things that Jervis says *might* happen *will* happen. But even setting aside the propriety of thus assuming that one knows the uniform underlying causes of human

behavior in given situations, the *ceteris paribus* clause renders such assertions useless as policy advice. Take the more familiar case of the minimum wage. The mere claim that, *ceteris paribus*, an increase of $X per hour in the legal minimum *must* cause more unemployment is wrong: If the increase is $.01, it may cause no unemployment at all. Nor can we know the level at which this will no longer be the case. So a *ceteris paribus* clause will, for policy purposes, at least have to be translated into a "tendency" for the wage increase to cause unemployment. Yet if the tendency may amount to zero this prediction is useless, and if the tendency translates into one or a hundred or a thousand people unemployed, policy makers may think that this cost is outweighed by the benefit of higher wages for the remaining low-wage workers.

A policy prediction must therefore have a quantity attached to it if it is to be of any service. Here economists oblige by conducting empirical research, but, of necessity, economists' studies focus on past times and places, often producing widely varying results—not surprisingly, if we do not assume that people everywhere and always will have the same *ideas* about how to respond to a regulation such as a minimum-wage increase (or a double-hull requirement).[17] Economists often proceed as if the variations in their empirical results can be ignored in favor of concluding that the "weight of the evidence" reveals that, say, minimum wages tend (or do not tend) to increase unemployment, but again, this presumes what is at issue—whether *knowable* (indeed, in this case, *known*) underlying laws are at work—and it does not produce a quantitative estimate.[18] To produce such an estimate, the economist as "policy scientist" must combine the results of past research into formulae that, like election predictions, will yield precise numbers—but only by, again, begging the epistemological question. The question is whether observers can reliably predict the behavioral reactions that policy actions will cause. Even if these reactions were based on the same unambiguous objective conditions that might explain data from the past, there would remain the problem of predicting what these conditions will be in the future. But if the people who react to the policy had and will have varied perceptions of the new situation caused by the policy, and divergent theories about how to respond to it, then unless we can read their minds, our point predictions will not be likely to hit their mark.

In short, an economic policy "science" is precisely as scientific as clairvoyant telepathy is. This is true of any other would-be policy

science, too, where changes in human behavior are the aim of the policy. Does this lead to the conclusion that we should not make public policy?

The question is nonsensical, as "inaction" cannot be logically distinguished from "action." The only choice is between various predictions of system effects: Will a particular *new* "state" action produce better results than would the mere continuation of the whole ensemble of previously enacted regulations (e.g., legal enforcement of a certain bundle of private-property rights), or would refraining from the new action produce better results? Given the hazards of prediction discussed by Jervis, it is tempting to throw up our hands and say that we cannot possibly know. This is the dilemma discussed by Posner (2012) and by Philip E. Tetlock, Michael C. Horowitz, and Richard Herrmann (2012) below.

Jervis's solution, in his contribution to the symposium (Jervis 2012), is more pessimistic than it was in *System Effects*, which ended with a catalogue of general recommendations for making public policy more robust against system effects. Jervis concluded that the possibility of these effects does not entail

> that reforms must fail or that directed change is impossible, but that the game does not end after one or two plays and that new measures will be needed to cope with the new problems. So in criticizing the quotation with which I began this book, I do not mean to imply that it is a mistake to require tankers to have double hulls, but only that doing so would have multiple consequences, some of which could defeat the purpose unless other actions are taken. They can be, however: Special instructions and training could be given to ships' captains, additional taxes might be levied on pipelines, and officials could be ready to respond to the undesired consequences of these supplementary policies. (Jervis 1997, 294)

Now, however, he writes that "although I closed my book with a discussion of how understanding system effects can lead actors to take advantage of them, I would not want to claim that this is always possible. We should always ask of an action, 'What will follow, and how will we and others react and change?' But we should also realize the limits to our ability to answer, or at least to do so correctly" (Jervis 2012, 412).

Perhaps Jervis came to realize that fixes such as those he had suggested as add-ons to the double-hull mandate cannot be known in advance to work, any more than we can know whether the fixes would respond to actual problems caused by the mandate: Jervis was, after all, at least as I interpret him, not providing us with ironclad predictions of the negative

unintended consequences such a regulation would produce, but instead with a list of *possible* unintended consequences. To be more schematic, these possibilities, drawn from the application of neoclassical economic theory to the policy advocated by Earle, are known (to Jervis) as unknowns. (To Earle, they seem to have been unknown unknowns.)[19] Yet even if we know of ten possible negative unintended consequences of a given policy, we cannot know which of these effects will justify a policy response unless we can predict whether they are not mere logical possibilities but ''tendencies'' and, if so, what their magnitude is likely to be.

If economics provided the best account of unintended consequences, we would have to leave the matter there and simply recommend that policy makers be as sensitive as possible to the known unknowns identifiable by thinking about the incentives new policies might create. However, Jervis's books provide the basis for a better account: a theory of unintended consequences grounded in epistemic failure, not misaligned incentives. Although Jervis often invokes the incentive of actors to circumvent new regulations (Jervis 1997, ch. 2), his analysis does not *rely* on incentives: ''The results of actions are often unintended and . . . regulations often misfire,'' he writes, because ''actors can rarely be fully constrained and will react in ways that those who seek to influence them are unlikely to foresee or desire'' (ibid., 91). As we have seen in the cases drawn from foreign policy, this unpredictability need not depend on the incentive to outwit an opponent; Japanese and U.S. decision makers had every incentive not to misperceive each other's intentions, yet they did. Their acts of misinterpretation were unintended consequences of each other's previous actions and led to further, disastrous unintended consequences. The bad news is that this makes the task of prediction even harder than it is when we are trying to select the likeliest from a group of known unknowns generated by economic theory. How can we possibly predict the likelihood of unknown unknowns?

If the general source of unintended consequences is epistemological, not motivational, it is hard to imagine how we could try to predict them case by case (policy by policy), but we might be able to get some traction if we move to a more general level, asking fundamental questions such as whether unknown unknowns might, overall, be positive rather than negative: surely no law dictates that mistaken beliefs cannot lead to serendipitous results. Then there is the question of degree. For expository purposes I have presupposed a binary distinction

between the predictability and the unpredictability of ideas, hence the unpredictability of behavior; this is clearly an exaggeration. We are not completely opaque to each other, and in consequence, evidence of each other's thinking is not uniformly ambiguous. If we could pin down the sources of our occasional interpersonal transparency, we might be able to make suggestions about institutions that might capitalize on them. We might also be able to discern which types of problem-solving are likely to be vulnerable to negative rather than positive unintended consequences. Tetlock is engaged in a sweeping long-term study of effective prediction strategies, based on his disturbing findings about the errors made by experts trying to predict relatively simple future events (Tetlock 2005). The approach I am suggesting would be even more intellectually ambitious, since the causes of interpersonal opacity (and transparency) may be related to such factors as biological and cultural evolution; and the institutional upshots of such factors would have to take into account, on the one hand, the differences between contemporary human relationships and institutions and their biological and cultural predecessors and, on the other hand, the potential directions in which fallible, ideational beings might be advised to go.

Jervis's contribution, then, may not fit into disciplinary pigeonholes because it points to a type of grand political theorizing—theorizing about the human condition—that would be appropriate only to normative political theory. And even in that subfield, such theorizing has long been out of style.

## NOTES

1. On Converse, see *Critical Review* 18, nos. 1–3 (2006), republished as Friedman and Friedman 2012a; on Tulis, *Critical Review* 19, nos. 2–3 (2007), republished as Friedman and Friedman 2012b; on Tetlock, *Critical Review* 22, no. 4 (2010).
2. Mitchell 2009 is an excellent guidebook to complex-systems theory, not least in that the author frequently asks whether the theory is applicable to human realities, and does not hesitate to answer in the negative.
3. Sometimes, to be sure, he imports law-like generalizations from other disciplines, such as psychology, writing, for example, that "people tend to think that good things (and bad things) go together and thereby minimize the perceived trade-offs among desired values" (Jervis 1997, 230) and then citing a slew of psychology-journal articles. (Jervis 1976 is an extended engagement with the cognitive-psychology literature aimed at producing generalizations about the basis of misperceptions.) However, if there is one field that, in principle, *might* impose regularities on what would otherwise be the kaleidoscopic flux of

ideational possibilities, it would be psychology, since the members of a species might well share cognitive traits that constrain their ideas. This is not to say, however, that the reading I am giving would be endorsed by Jervis. I am trying to ferret out the logical prerequisites and implications of system effects as he presents them, but sometimes readers of the book will find statements, particularly in chapter 1, that run counter to my interpretation of this logic, particularly on pages 22 and 144, where Jervis casually refers to the "laws" of economics and, on page 22, of politics. (However, Jervis does not identify these laws or claim to be adding to them.) I construe this as a case of his not having fully appreciated the radically antipositivist implications of the book.

4. It does not take a leap of the imagination, however, to "predict" that political scientists might jump on the Nate Silver bandwagon and use statistics to try to predict things other than elections. See Ward and Metternich 2012. For a preemptive, measured critique of the applicability of statistics to future events, see Blyth 2006; and for discussions of the related work of Nassim Taleb, see Blyth 2009, Jervis 2009, and Runde 2009.
5. Usually the models "cheat" by using survey data on presidential approval, survey data on the two candidates before and after their conventions, primary-election results, and other measures of voters' opinions about the presidential candidates, which are variants of the very thing expressed by their votes on Election Day. We already have plenty of polls, however; if the forecasting exercises have a scholarly purpose, it is to show that "real" factors, such as changes in unemployment rates and economic growth in a given quarter prior to the election, are (somehow) at work, even though the real factors alone cannot make accurate predictions, and so must be tweaked by using polls and other direct measures of opinion. The forecasters then fit various measures of public opinion and real factors against the small N of past presidential elections to produce a model that will forecast the next one, which presupposes that there is a temporally uniform underlying mélange of causes expressible in a formula weighting the various factors. The notion seems to be that as the N grows over time, the formulae will grow more precise overall, and that if they fail in a given case, that is only to be expected, as these are mere probability predictions. In short, only the inapplicability of the *ceteris paribus* clause, not the inapplicability of the *Homo economicus* model itself, is considered as a cause of outliers. See Campbell 2012 and 2013.
6. An interesting exception is explained in Lewis-Beck and Tien 2013. The authors' best-performing model in 2012 used only a measure of subjective perceptions of objective economic conditions: namely, the net proportion of survey respondents six months before the election who said that "business conditions are worse" than they had been previously. The authors explain that on the basis of "voter behavior theory," they would have preferred their old "Jobs Model" to the subjective model, but the predictions of the two models diverged, with the subjective model performing better (ibid., 39). That is, people's perceptions of economic conditions were more predictive of how they would vote than were the real conditions themselves. If so, however, then perhaps the dependent variable that should be of interest is not the electoral outcome (which we all find out, soon enough), but people's perceptions of

economic conditions; and perhaps the hypothesized independent variables should be factors (such as media coverage of economic conditions) that could produce these perceptions regardless of whether they accurately reflect real conditions. The election-forecasting scholarship is pointless curve fitting unless the models are supposed to identify what is *really* behind people's votes. But if this "real" factor is people's beliefs, then the economic factors used in the likes of the Jobs Model should be seen as, at best, proxies for what is actually causal: voter perceptions (whether of unemployment, business conditions, or anything else).

7. E.g., Fiorina 1981.
8. See n6 above.
9. Or they get washed out by the use of measures of opinion to diminish the impact of real factors alone; see n5 above.
10. Moreover, even if one votes out of a felt civic obligation rather than as an attempt to affect the outcome, this obligation is not fulfilled merely by voting per se. Few would claim that they have a civic obligation to show up at the polls but that, having done so, they may proceed to choose whom to vote for by flipping a coin. (How would such an obligation make sense?) Nor would an instrumentally rational voter who thought her vote was likely to be decisive have reason to try to affect the outcome (by voting) if she could not motivate a nonrandom vote. An obligation to vote, or a desire to affect the outcome, must entail voting for the "right" candidate, i.e., the one who, the voter predicts, is likely to advance what she takes to be good ends. Yet if people *knew*, as rational-ignorance theory holds, that they were too poorly informed to make such predictions with any reliability, because they had *deliberately* underinformed themselves, then they would have no reason, whether moral or instrumental, to vote.
11. A rejection of the pretense of predictive knowledge does not entail a rejection of determinism. I am suggesting that ideas are causes of behavior. In a Laplacean sense, one could, *in principle*, predict ideas, hence behavior, if one had omniscient command of all the antecedent conditions that lead one person to invent or endorse or transmit an idea, while another rejects it or never hears of it in the first place. However, we do not have such knowledge, and it is a safe bet that we never will. A predictive epistemology is logically possible but pragmatically impossible. The "laws" of epistemology may be knowable in principle, but not in practice.
12. Another path toward minimizing the role of subjective beliefs in human behavior is to treat emotions as overriding (rather than being triggered by) subjective beliefs, since emotions can be assumed to have some roughly general similarity across individuals, and these similarities can plausibly be seen as objective facts that directly control individual behavior. Long ago, Jervis (1976, 4–5) dispatched political psychologists' overemphasis on emotions by pointing out the performative contradiction it involves, at least if one is using emotion to explain the behavior of policy makers in non-crisis situations. The scholars attributing agents' behavior to emotion surely would not attribute their own attribution of emotion to the agents as the result of emotion. Cf. Friedman 2012 and Ross 2012.
13. This is not to say that there are no differences over values, any more than by bracketing the role of emotion, one is saying that emotion never overrides, rather than amplifying, rational judgment. But if one does not set values and

emotions to the side, one cannot even consider the possibility of subjective misperceptions of objective facts.

14. On complex-systems theories as theories about epistemology, not ontology, see McIntyre 1998.
15. Clearly, in this respect, I am departing from the letter of Jervis's book, but I hope not from the spirit.
16. I am referring to Hayek's quantitative notion of what makes "spontaneous" orders "complex phenomena." See Hayek 1967.
17. See Neumark and Wascher 2009 for summaries of many dozens of studies of the effects of minimum-wage increases.
18. In response to White House claims that "a range of economic studies show that modestly raising the minimum wage increases earnings and reduces poverty without measurably reducing employment," a *Wall Street Journal* editorial quoted David Neumark of the University of California at Irvine, who said that "the White House claim of de minimis job losses 'grossly misstates the weight of the evidence.' About 85 percent of the studies 'find a negative employment effect on low-skilled workers.'" "The Minority Youth Unemployment Act," *Wall Street Journal*, 19 February 2013.
19. Or, at best, putative known unknowns whose applicability Earle had reason to doubt.

## REFERENCES

Blyth, Mark. 2006. "Great Punctuations: Prediction, Randomness, and the Evolution of Comparative Political Science." *American Political Science Review* 100(4): 493–98.

Blyth, Mark. 2009. "Coping with the Black Swan: The Unsettling World of Nassim Taleb." *Critical Review* 21(4): 447–65.

Campbell, James E. 2012. "Forecasting the 2012 American National Elections." *PS: Political Science and Politics* 45(4): 610–13.

Campbell, James E., ed., 2013. "Recap: Forecasting the 2012 Election." *PS: Political Science and Politics* 46(1): 37–48.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Friedman, Jeffrey. 2012. "Motivated Skepticism or Inevitable Conviction? Dogmatism in the Study of Politics." *Critical Review* 24(2): 131–55.

Friedman, Jeffrey, and Shterna Friedman, eds., 2012a. *The Nature of Belief Systems Reconsidered*. London: Routledge.

Friedman, Jeffrey, and Shterna Friedman, eds., 2012b. *Rethinking the Rhetorical Presidency*. London: Routledge.

Friedman, Jeffrey, and Shterna Friedman, eds., 2013. *Political Knowledge*, 4 vols. London: Routledge.

Friedman, Milton. 1953. "The Methodology of Positive Economics." In idem, *Essays on Positive Economics*. Chicago: University of Chicago Press.

Hayek, F. A. 1967. "The Theory of Complex Phenomena." In idem, *Studies in Philosophy, Politics, and Economics*. Chicago: University of Chicago Press.

Hetherington, Marc J. 1996. "The Media's Role in Forming Voters' National Economic Evaluations in 1992." *American Journal of Political Science* 40(2): 372–95. Republished in Friedman and Friedman 2013, vol. 3.

Holbrook, Thomas, and James C. Garand. 1996. "Homo Economicus? Economic Information and Economic Voting." *Political Research Quarterly* 49(2) (June): 351–75. Republished in Friedman and Friedman 2013, vol. 3.

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton: Princeton University Press.

Jervis, Robert. 1997. *System Effects: Complexity in Political and Social Life*. Princeton: Princeton University Press.

Jervis, Robert. 2009. "Black Swans in Politics." *Critical Review* 21(4): 475–89.

Jervis, Robert. 2012. "System Effects Revisited." *Critical Review* 24(3): 393–415.

Jones-Rooy, Andrea, and Scott E. Page. 2012. "The Complexity of System Effects." *Critical Review* 24(3): 313–42.

Lewis-Beck, Michael S., and Charles Tien. 2013. "Proxy Forecasts: A Working Strategy." In Campbell 2013.

McIntyre, Lee. 1998. "Complexity: A Philosopher's Reflections." *Complexity* 3(6): 26–32.

Mitchell, Melanie. 2009. *Complexity: A Guided Tour*. New York: Oxford University Press.

Monteiro, Nuno. 2012. "We Can Never Study Merely One Thing: Reflections on Systems Thinking and IR." *Critical Review* 24(3): 343–66.

Monteiro, Nuno P., and Keven G. Ruby. 2009. "IR and the False Promise of Philosophical Foundations." *International Theory* 1(1): 15–48.

Neumark, David, and William L. Wascher. 2009. *Minimum Wages*. Cambridge, Mass: MIT Press.

Page, Scott E. 2011. *Diversity and Complexity*. Princeton: Princeton University Press.

Posner, Richard A. 2012. "Jervis on Complexity Theory." *Critical Review* 24(3): 367–73.

Ross, Lee. 2012. "Reflections on Biased Assimilation and Belief Polarization." *Critical Review* 24(2): 233–45.

Runde, Jochen. 2009. "Dissecting the Black Swan." *Critical Review* 21(4): 491–505.

Tetlock, Philip E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press.

Tetlock, Philip E., Michael C. Horowitz, and Richard Herrmann. 2012. "Should 'Systems Thinkers' Accept the Limits on Political Forecasting or Push the Limits?" *Critical Review* 24(3): 375–91.

Tulis, Jeffrey. 1987. *The Rhetorical Presidency*. Princeton: Princeton University Press.

Ward, Michael D., and Nils Metternich. 2012. "Predicting the Future Is Easier than It Looks." *ForeignPolicy.com*, 23 November.

Weber, Max. [1904] 1949. "'Objectivity' in the Social Sciences." In idem, *The Methodology of the Social Sciences*, trans. Edward A. Shils and Henry A. Finch. New York: Free Press.

White, Wallace. 1989. "Her Deepness." *New Yorker*, 3 July: 41–65.