

# U.S. Foreign Policy: Theory and Explanation

**Branislav L. Slantchev**

*Department of Political Science, University of California, San Diego*

Last updated: January 4, 2016

In everyday use, the word **theory** has very unfortunate connotations. It is often taken as equivalent to *speculation* or *idea*, as in “human-caused climate change and evolution are just theories” (with the implication that they lack the empirical evidence necessary to support their claims), or to **hypothesis**, as in “I have a theory where my missing car keys might be” (with the implication that it is a conjecture that is subject to empirical verification). In science, however, the word **theory** refers to a well-articulated *explanation* of some phenomena that is also *well-supported* by empirical evidence. In this sense, it is well beyond mere speculation: it tells us *how* the phenomenon in question works, and this explanation generates a variety of testable propositions, or *hypotheses*, that can be, and have been, evaluated empirically through experiments and observation. The overwhelming evidence supporting these hypotheses is then taken to indicate support for the theory itself. Although no theory is ever “proven” in the sense of a mathematical theorem, at some point the evidence in support is so overwhelming that only an insane person who is made aware of it would deny it. Heliocentrism is “just” a theory, but I dare you to find any reasonably educated person who would deny that the Earth orbits around the Sun.<sup>1</sup>

The opposite of theory is not fact but mystery.

## 1 Causal Mechanisms

Our use of the word *theory* is not going to be anywhere near the scientific ideal but it will be much more demanding than the everyday use. For us, a theory must provide an explanation: a **causal mechanism** that tells us how some variables interact with each other to produce the outcomes we seek to understand. Notice that the selection of theory depends on the target: what is the question we seek to answer? The question is usually something that confounds our expectations, something that we do not understand, and so something that needs to be explained. Theory provides the answer in the form of a mechanism that establishes a causal chain between the variables and the outcome.<sup>2</sup>

---

<sup>1</sup>It is astounding that in 2012 one in four Americans still believed that the Sun goes around the Earth. See Table 7-8 in the report by the National Science Foundation, <http://www.nsf.gov/statistics/seind14/content/chapter-7/c07.pdf>. Europeans fared even worse: in 2005 one in three failed this fundamental astronomy question. I have no data on the distribution of incorrect answers by educational level but it is stunning that one can get out of high school and still believe the equivalent of the Earth being flat and resting on the back of a world turtle, with larger and larger turtles all the way down.

<sup>2</sup>In this way our theories are not merely instances of abstract generalized thinking, as in “music theory” or “art theory” or “literary theory”, but instead specify models/mechanisms that ultimately yield testable proposi-

Let us start with some historical phenomenon that we might wish to understand. The most obvious problem with history is that there are too many variables one could potentially look at. Which are important and which can safely be discarded? How do we decide? The answer is that we need a “guide” to selecting variables. This is what theory does: it tells us how some variables interact with each other to produce the phenomenon in question (which is another way of justifying the need to look at these, and not other, variables).

Consider a hypothetical example. Suppose we observe a statistical correlation between war initiation and high unemployment. Our hypothesis would be that high unemployment causes wars of aggression. We now need a theory that provides the mechanism that links the explanatory variable (unemployment) to the explanandum (war of aggression). We can hypothesize that high unemployment (a) causes social unrest that could be channeled toward an enemy, (b) causes governments to expand employment in armament industry — reduces unemployment and is justified by attributing hostile intent to enemy, (c) causes governments to search new markets to encourage producers to hire workers — aggressive foreign policy, (d) gives rise to populist leaders who are more aggressive in foreign policy. We could now use this theory to check whether the cause has the hypothesized effects which in turn produce aggressive wars. But we could also continue to refine the theory by opening up (d): why would high unemployment bring populist leaders to power? We could theorize that (d-1) the natural clientele of populist is more likely to vote (or engage in political behavior) when its opportunity costs are low — which they will be when unemployed since there is no income to forego; (d-2) populists are more likely to promise instant solutions to unemployment; (d-3) populists offer to punish those that the unemployed believe to be responsible for their plight. Again, each of these hypothesized effects can be checked against data. But we do not have to stop there: we could want to know how those “guilty” for the plight of the unemployed are identified and punished. We might hypothesize that (d-3-1) the wealthy would be worried about the security of property rights and so would be willing to strike deals with the government in which they relinquish some of their wealth in return for protection — redistribution toward the unemployed; (d-3-2) they might support the leader in aggressive foreign policies that blame the enemy in an effort to deflect attention from themselves. These hypothesized effects would predict that high unemployment would be associated with some internal redistribution of wealth and with propaganda vilifying an external enemy. The latter can lead to crisis escalation and, possibly, war.

## 2 Rationalist Explanations

For a mechanism to be of any use, it has to go beyond providing a list of variables and effects. Since the phenomenon we are interested in here (war) is ultimately produced by the behavior of people, a mechanism should be anchored in individual behavior. In other words, it should tell us why the relevant agents acted in particular ways in given contexts. But how do we understand individual behavior — generally, we do so by **rationalizing** it. That is, we take the observed behavior we seek to understand, and then attribute some preferences and beliefs to the individual that engaged in it such that this observed behavior is expected to contribute to the welfare of that individual as defined by his beliefs and preferences.

---

tions (hypotheses) that can be subjected to experimental and quasi-experimental evaluation.

We assume that individuals are “rational” in the sense that their actions are purpose-driven so that individuals tend to behave in ways that are supposed to enhance their well-being. How individuals define well-being and how they analyze their environment depends on their preferences and beliefs. The actions they can choose from depend on the context in which they act and the information they have; that is, on institutional and informational constraints. An idealized “rational agent” always chooses the optimal course of action, with “optimal” defined as the course most likely to deliver on the desired goals.

All of this is purely hypothetical: we use observed behavior to infer preferences and beliefs that make this behavior optimal given the constraints. We then explain the behavior by saying that it must have been the result of the purposeful pursuit of the goals we attributed to the individual. This sounds suspiciously ungrounded in reality, and it would be without some means of testing the various connections this mechanism requires in order to make the causal chain work. The virtue of having the theory is that it tells one which variables to look at, how they should change, and what their effects should be — all of this can be subjected to empirical testing (observational or experimental). We could attempt to ascertain the preferences and beliefs the relevant individuals had to see how closely they match our assumptions about them. We can go further and ask whether it is *reasonable* for the individual to have held these beliefs given the information this individual had at the time. We would also attempt to analyze how closely the constraints we assumed are matched by the context in which the individual had to act. Matching closely these factors would give us confidence that the mechanism we postulated is, in fact, explaining behavior. We could say that we understand it because we can rationalize the behavior of the relevant individual with some confidence.

Why focus on rationalist explanations? For starters, people *want* to be rational in the sense we’ve been using the word. They want to have “good” reasons for their behavior, which is why they often “rationalize” them after the fact by pretending to have had goals or beliefs that would make their behavior reasonable. More importantly, we rely on this sort of reasoning all the time when we want to make sense of the behavior of others and when we want to predict how others will react. In fact, when we fail in these predictions we are apt to characterize the surprising behavior as irrational.

This is not to say that “irrational” behavior must be unintelligible. For example, strong emotions might short-circuit decision-making and cause individuals to rush into actions they otherwise would not have. Shame might cause one to commit suicide; fear might cause another to jump out of a burning building. Desire for revenge might motivate actions that are exceedingly costly personally with little objective benefit even if they succeed. (In these, however, some element of ratiocination might remain if the individual still chooses the course of action that is most likely to cause the desired result.) Weakness of will is often behind failure to lose weight or, in some cases, quit smoking. Wishful desires bias belief formation, causing individuals to stop searching for better solutions or more information, or to discard information contrary to their desires. There are many other psychologically motivated biases in decision-making that might produce actions that fall short of the optimal. Going into psychiatric explanations, there are also the various obsessions, phobias, delusions, and so on. Any of these can make behavior intelligible, so why should we privilege rationalist explanations?

The main reason for that is that irrationality can “explain” too much too easily. People

often attribute puzzling behavior to irrationality when in fact it could be perfectly rationalizable by factors they fail to consider. Take, for example, the Marxist hypothesis about *false consciousness*. According to Marxism, the proletariat does not have a shared interest with the capitalists in policies that enhance the well-being of the latter (because this could only increase the exploitation of the former). An example of such a policy, at least according to Lenin's view, would be "imperialist wars," that is, wars fought by capitalist societies over access to markets and colonies for raw materials. Since it is precisely the members of the proletariat who die as soldiers in these wars but only the capitalists stand to reap the profits, it is in the workers' interest not to support such wars. When the First World War broke out, many Marxists in fact expected the masses to recoil from service. Unfortunately (for theory and for the masses), the opposite happened — not only did proles from one country enlist in their armies, in many cases voluntarily, but they did not seem particularly reluctant to kill "fellow" proles from other countries with whom they supposedly shared interests in overthrowing capitalists. This was a clear divergence from behavior that class interest would dictate. The theory was "saved" by the notion of "false consciousness" according to which the ideological control of society by the bourgeoisie and nobility has blinded the proletariat to its true class interests. The proles either do not know that interest (because, for example, religion tells them what the "natural order of things is") or they do but choose to disregard it because they are promised to enter the ranks of the privileged. Whatever the reason, the proletariat's acting against the interests postulated by the theory is "explained" by amending the theory to essentially argue that the proletariat is deluded. (A much simpler explanation would have been that the theory is wrong.) Thus, according to Marxist theory, the proletariat will act in its own interest except when it does not. Observationally, when we observe workers unionizing and striking, the theory is supported because it is in the interests of workers to force the capitalists to share in the surplus their labor creates. When we observe workers acting in concert with capitalists to thrash other workers and their capitalists, the theory is supported because they are acting out of false consciousness.

There is no possible behavior that the workers can engage in that can *falsify* the theory, even in principle. This means that we have to take the theory on faith — there simply exists no sort of evidence that could potentially disprove it. But if the theory were wrong, how would we then know this? In the above example, we could not. This renders the theory useless as an explanatory device: everything that does not conform to one postulate conforms to another in the same theory. We shall require our theories to have a property known as **falsifiability** — meaning that if the theory is false, then there does exist some sort of evidence we can obtain either by observation or by experiment that would demonstrate that. Without false consciousness, Marxism is falsifiable — the evidence of workers failing to act in their class interests would show that the theory is wrong. With false consciousness, Marxism is unfalsifiable since all evidence is consistent with the theory. It is not that one should discard a theory at the first sign of non-conforming evidence — that would be naïve. One can always seek to amend the theory to account for any new evidence in addition to all the evidence it could previously handle. However, when such an amendment goes too far — like false consciousness does — it can render the resulting theory unusable.

Rationalist explanations are in a way *minimalist explanations* because they are the ones most readily falsifiable. This makes them particularly suitable for hypothesis testing, which allows for accumulation of knowledge and verification. Explanations that rely on irrational-

ity do not have to be non-falsifiable (although some of them are). The problem is that they are too convenient and so might lead to ignoring the actual mechanism. It is all too easy to say “oh well, he acted out in anger” instead of searching for other causes explaining puzzling disregard for one’s own safety. In fact, the ability to mimic irrational behavior for rational reasons should give one further pause before reaching for such explanations. If an individual “acts crazy” for the purpose of convincing others that he is crazy (meaning that they cannot rely on usual cost-benefit reasoning to predict how he would act), he is not really crazy — provided the others believe him and adjust their behavior accordingly. He is cunning, he is strategic, he is supremely rational in choice of action given his goal.

To give a specific example, how are we to understand the 2003 Iraq War or, more specifically, how are we to understand the behavior of Saddam Hussein? In the light of the outcome of the 1991 war over Kuwait, the subsequent degradation of the Iraqi armed forces, and the continued improvement of the US military, it would appear nearly certain that a war with the US would have inevitably ended in the overthrow of the Iraqi dictator. So why pursue policies that clearly tilted the US toward war and, more importantly, why persist after it became clear that the US will, in fact, invade? One answer is that Hussein was irrational, so these calculations simply did not enter his mind. He might have put his faith in God or in his own genius. This, however, sounds more like a label than an explanation. One could instead argue that Hussein made a mistake because he was misled as to the true state of his military by advisors who were too afraid of him to reveal just how much it had deteriorated. This would have given him false optimism and encouraged him to resist. (Similarly, he might have expected the US to be incapable of forming a grand Coalition of the 1991 type — which was correct — and thus be reluctant to fight on its own — which was incorrect.) This explanation would rationalize his behavior by showing that it was reasonable given the information he had at the time. An even stronger version would argue that even while there was no uncertainty about the military outcome of an American invasion, there was far more uncertainty as to the fate of subsequent pacification — would the Americans have the stomach to stay and fight for years on end an enemy that mingles with civilians and that cannot be readily identified and defeated in pitched battle? If Hussein could survive the initial onslaught and then organize national resistance to the occupying forces, then resisting the US makes sense especially if failure to do so would expose the weakness of the dictatorship and make Hussein’s overthrow nearly certain. This type of explanation rationalizes his behavior by showing that he took a calculated risk, a risk that actually made sense despite the overwhelming military superiority of the United States. Even though he eventually failed, the behavior had been reasonable. Which of these (or the myriad alternative) explanations is valid depends on the assessment of the facts and how closely they track the connections identified by the various theoretical mechanisms.

### **3 The Map Analogy**

A final word about theory: it is *not* a full description of reality. It cannot be: the closer it gets to reality the less useful it becomes as a means of understanding that reality. The power of theory is in that it abstracts away from the complex real world and attempts to reduce its vastly complicated interrelationships to a small set of manageable variables and connections. In this, a theory is like a map. How useful this simplification is depends on

the purpose (which determines how much detail you can omit without producing a useless map) and how good the theory is (it includes all the variables it has to in order to produce reliable predictions about their effects). Neither of these is really known *a priori*, so each theory is essentially a bet that its particular formulation would be useful.

Each theory is then “valid” while it continues to be useful. It is not discarded when one encounters contradictory evidence, especially if there exists no alternative that can take its place. The theory can be modified to account for that new evidence although care should be taken that the adjustment is not ad hoc, meaning that the new version should handle what the old theory could plus the new evidence plus whatever new hypotheses it gives rise to. It is a tough order for a new theory to pass, which is why we have long used theories known to have “holes” in them — Newtonian physics is one example, Ptolemaic astronomy is another — they are good enough for most purposes and there was no viable alternative — until, that is, Einstein’s theory of relativity and Copernicus’ theory of Heliocentrism.

Going back to our map analogy: how useful would it be to have a map that is an exact representation of reality? For starters, it would be impossible to create one: it would have to be as large as the world it represents. OK, so the first “compromise” would be to reduce it to manageable proportions, say 1 to 5,000 (1 cm to 50 m), which would be useful for a walking map. Obviously, going that small means discarding a lot of detail. So what can we let go? It depends on the purpose of the map. If we want a walking map, then we should retain roads, paths, trails, some information about the terrain, and relevant markers. If we want a driving map, we need roads but can omit foot trails, we might want to include gas stations and rest stops, and so on. A walking map would not be useful in a city if we wish to use the bus, and a map of the bus routes would not be useful if we need to use the subway. In fact, anyone who’s ever looked at a map of bus routes or subway lines would be familiar with the highly idealized schematic representation of reality they represent — nice straight lines with nice junctions at right angles and often stations equidistant from each other — in short, very little of reality has made it onto these maps. Yet they are far more useful for those trying to utilize the respective modes of transportation than a highly detailed physical map of the place or a nicely illustrated map of tourist attractions.

Theories work the same way: purpose determines scale and simplification. The trouble is that unlike a map — where purpose fairly clearly dictates content — no such useful guide exists for theories. We have to formulate them, produce tentative hypotheses, proceed to experimental and observational verification, then re-formulate as necessary. No theory is ever final (and that’s a good thing) — theories are always the best we can do with the knowledge we currently have. This makes them tentative and subject to revisions. Theories that have withstood the test of time acquire the special status of scientific “truth” because we have yet to uncover disconfirming evidence. But this “truth” is not absolute, it is not dogma. It is no more nor less than a reflection of what’s possible in our state of the world.