Anarchy Is What Individuals Make of It

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Andrew Justin McKenzie, B.A.

Graduate Program in Political Science

The Ohio State University

2013

Master's Examination Committee:

Randall Schweller, Advisor

Alexander Wendt

Abstract

Theories and models of political behavior, while sometimes predicated on methodological individualism, routinely fail to consider the possibility and potential impacts of human free will—or the implications if humans lack free will. I argue that all models of social behavior, whether individualistic or holistic, must take at least an implicit position on whether individuals can make free (i.e., autonomous) cognitive and behavioral choices. However, social scientists' everyday agnosticism on the question of free will threatens theoretical falsehood and practical irrelevance. I discuss the consequences for political science—focusing on international relations—of the existence or absence of free will. I use metapreferences as a modeling technique to help us conceptualize how free will and causation interrelate, and from this develop the argument that free will elevates the importance that natural science and technology play in creating preferred social outcomes. I close by applying the preceding arguments to the study of war.

Vita

2005 …… B.A. Diplomacy & Foreign Affairs, Miami University

Fields of Study

Major Field: Political Science

Table of Contents

List of Tables

Chapter 1: Introduction

On December 17, 2010, a twenty-six-year-old Tunisian produce vendor named Mohamed Bouazizi was confronted by a local policewoman for operating an unlicensed produce cart.[1] For Bouazizi, the confrontation was the last straw in a series of harassments and perceived mistreatment. He resisted the policewoman and was allegedly beaten in response by the policewoman and her colleagues.

Humiliated, Bouazizi went to the local municipal offices to complain. He was ignored and eventually left, but returned to the plaza in front of the municipal building an hour later with a can of paint thinner. There, he ranted against his mistreatment and government indifference, doused himself with the paint thinner, and ignited it. He died from his burns a few weeks later.

Local outrage at the events leading to Bouazizi's self-immolation soon expanded into widespread anti-government protests that spread across the country. Frustrated with unemployment and economic malaise, political inefficacy, bureaucratic corruption, and police brutality, the protests only grew in response to official efforts to crack down. Tunisian president Ben Ali's moves to publicly assuage dissidents while privately increasing suppression backfired, and the government fell.

---

[1] For background, see Abouzeid (2011), Anon. (2010, 2011), Byrne (2011), Fahim (2011), Rohr (2011), Ryan (2011a, 2011b), Whitaker (2010), and Worth (2011). A number of the minor details, including the exact nature of the confrontation, are inconsistent across accounts.

[2] See Bigo (2013).

[3] See McCarty and Meirowitz (2007) on "thick" and "thin" models of rationality and Bueno de Mesquita

In the meantime, protests had spread across the Arab world, with neighboring Libya and Egypt's leaders and governments both challenged and deposed, Syria amid a multi-year civil war, and even the 2012 U.S. Presidential election focusing, among other things, on the incumbent president's handling of the "Arab Spring." While the long-term consequences, domestically and internationally, of the Arab Spring are still unclear, few could claim with much warrant that the revolutions are historically meaningless or unimportant.

Whether, factually and historically speaking, Bouazizi's suicide led to significant political outcomes is fairly indisputable. Of course, whether Bouazizi's atypical act is *theoretically* relevant in explaining the Arab spring is more open to discussion; after all, structuralist perspectives would contextualize what Bouazizi did in light of the preexisting social situation.[2] Nevertheless, as I will argue more completely in section two, such a contextualization raises questions along the lines of how such preexisting social situations are constituted by human actions, and whether these actions can be as atypical as Bouazizi's. Humans behave in a multitude of ways, and while there are clearly some behaviors that have empirically been more routine than others—for example, having children is a more regular part of most human experience than self-immolation—we must ask: what freedom of action does each human have, and how robust are political science models in light of our answer to this question?

As it stands, much of the time, political science models say virtually nothing about individuals and our freedom of behavior; we are apparently but pawns in a game

[2] See Bigo (2013).

played by markets, states, historical eras, and the like. Even when humans are the players of interest in models, human behavior is usually assumed to tend toward neatly defined end-points, or to follow well-worn psychological paths. Thus, even in "microfoundation"-heavy models, humans are—to borrow Hollis's (1983) term—nothing but "throughputs" who mechanistically process and execute exogenously given actions, preferences, and so forth.

What are political scientists to make of unexpected human behaviors, like Bouazizi's suicide, or the world-shocking self-immolation of Vietnamese monk Thich Quang Duc during Vietnam's civil war? Are they merely epiphenomena of existing structures and social games? Or might they be "black swans" that thwart widely accepted theories and their predictive power?

Puzzle: what if we're free?

The "puzzle" of this paper, then, originates when we consider human actions that do not appear, on the surface, to be either materially (in the sense of material self-interest) or socially determined, and to ask whether such "abnormal" behaviors should be *assumed* to always remain abnormal and irrelevant to political scientists. For instance, self-immolation seems to blatantly contradict the "thick" notion of self-interest and, as we will see later, even raises questions about the "thin" notion of self-interest.[3] After all, did Thich Quang Duc's immolation simply reveal a static preference or represent

[3] See McCarty and Meirowitz (2007) on "thick" and "thin" models of rationality and Bueno de Mesquita (2009) on the usefulness of conceptualizing "brazen self-interest"; these issues are explored further in section three.

changing prices, along the lines of Stigler and Becker (1977)? Is it useful or valid to think of self-immolation in a microeconomic framework?

But while many political scientists would probably claim to reject the reduction of Thich Quang Duc or Bouazizi's decision to self-immolate into a microeconomic model, political science theories consistently resort (perhaps of necessity—as will be discussed) to claiming that human decisions are, in essence, "made for us"' by social and institutional settings, historical circumstances, biological constraints, or predetermined neurological processes. In each case, whether microfoundations are present or not, humans end up looking like Hollis's throughputs, and one individual's decision to set himself on fire in protest looks no different than a congresswoman's decision to seek re-election or a businessman's decision to invest in a profitable new venture.

The question—and challenge—of this paper is whether by modeling decisions in such a "scientific" way we do not inevitably foreclose on an individual's autonomous inner world of deliberation, introspection, self-critique, and freedom of action, and thereby lose something important in the process: consistent with our models, we lose the ability to believe that individual decisions and social behavior could possibly go any other way than they do. Perhaps just as worrisome, this paper challenges that such thinking paints political scientists into an awkward corner: even if political scientists argue conclusively that humans lack the ability to make free choices and that, indeed, all human action is entirely predetermined, the question arises of what is predetermining our behavior: fundamental physical and chemical laws or political-scientific models?[4]

[4] What would Friedman say?

Therefore, this paper argues that neither a consistent belief in human free will nor a consistent belief in behavioral determinism is neatly compatible with political science as practiced. Therefore, the question of whether humans do indeed possess "free will"—as defined in the incompatabilist sense, to be discussed momentarily—should be a question of paramount importance to political scientists. Yet with a few mostly confused exceptions, free will is not discussed nor even, it would seem, professionally considered by the majority of practicing social scientists across all major traditions and paradigms.

In this paper, I will argue that the nature of the human will deserves a thorough examination, especially one drawing connections between the two dominant incompatibilist perspectives on free will and the implications of those perspectives for how—and why—social science is practiced.

The "unproblematization" of will in political science

To say that free will is not problematized in political science is not quite the same as saying say that free will is never brought up by political scientists. Some researchers, such as Walzer (2004), raise the topic of free will—only to claim that they will try to avoid reference to it, as if the notion of free will is yet another philosophical quagmire to be avoided. Meanwhile, writing in Science (as excerpted in the *Wilson Quarterly*), Shaun Nichols does look at free will—but only to recount results of social-scientific experiments investigating the "psychological mechanisms" undergirding participants' answers to questions about free will.

At the opposite end, Davies (2004) wants to tackle the topic of free will head-on. But Davies' concern is not how free will interfaces with international relations research, but rather a more practical worry, viz., that the idea that human behavior is predetermined "will be oversimplified and used to justify an anything-goes attitude to criminal activity, ethnic conflict, even genocide. … The [natural] scientific assault on free will would be less alarming if some new legal and ethical framework existed to take its place." (37)

Thus, although free will is examined in a sparse smattering of political science (or related) work—see, for prime examples, Williams and Mayer (1962), Glock (1964), Adcock (2007), Carr (1939), Doty (1997), Kurki (2006), and Wallerstein (1997), the more recent of which (as well as Carr) will be discussed later—overall, free will is not treated as a threat to political science as argued in this paper. Among major books in the field of international relations in the last few decades, for example, only Schweller (2006) even mentions free will—once, when he states in passing, "[U]nless one accepts a rigidly deterministic view of historical events, the notion of free will and human responsibility must apply to any actual state of human affairs and to human actions." (18) Otherwise, even researchers known for delving into philosophical hot-topics have been silent on the question of free will. Wight's (2013) situation of international relations theory amid the philosophy of social science does not bring up the topic, coming closest in his mention that "for post-structuralism there is no underlying logic to structures and hence there is structural indeterminacy," a point emphasized below (34).

Given that political science, and especially international relations scholarship, has a reputation for descending regularly into philosophical debates, why is it that one of the

most major of all philosophical debates is so routinely ignored by political scientists? What may be the biggest problem with analyzing free will in social science is that it seemingly requires theorists to give up on the idea of neat, straightforward causality and the ability to use external factors such as material incentives/imperatives, structures, and the like to "compel" individuals in society to behave according to what a theory predicts. This accords with Wendt's (1997) table depicting IR theories (and the surrounding discussion), which reveals a relative lack of theoretical activity in the category of individual idealism, in which theorists minimize material as well as holistic sources of causality (31–33).

Nevertheless, the oversight is puzzling, especially because of how close to the surface the philosophical debate lies to many relevant theoretical and empirical issues in international relations—for example, calling the payoff from defecting in the Prisoner's Dilemma game the "temptation" payoff, or referring to states as having "resolve" in the bargaining on the eve of war.

Do we have microfoundations?

To frame some of these questions another way: for decades a lively debate in social science has concerned calls for greater attention to the individual "microfoundations" (e.g., Friedman 1957, Lucas 1975, Achen 2002) underlying macro- or aggregate behavior. Indeed, one move in this paper is to reemphasize the importance of microfoundations in structural and structuration theories.

Yet despite a good deal of abstraction in microeconomic and other microbehavioral models, human individuals remain automatons, their behaviors and preferences neatly delimited or determined by the theorist. Thus, in contrast to the cry for microfoundations, many social scientists working congruent with methodological individualism nevertheless seem content to leave individuals' behavioral microfoundations themselves unfounded (e.g., Gerber and Green 2000), even when drawing on behavioral economics and cognitive psychology. (This is not to say that specific researchers or research projects can never exogenize/bracket factors and variables, only that making field-wise behavioral assumptions is a dangerous game.) Therefore, I argue that we should think more about the philosophical "microfoundations" of human action underlying these microeconomic/microbehavioral-styled microfoundations that support most individual-level thinking in the social sciences.

For example, Achen bemoans quick acceptance of good-fit empirical results that do not match up with a microfoundational model supporting the assumptions behind the statistical models leading to such results:

> [W]e have yet to give most of our new statistical procedures legitimate theoretical microfoundations …. Typically, no formal model supports [a particular estimator's] assumptions, and no close data analysis is presented in their favor. In fact, no matter how devastating those absences, we often write as if we didn't care. Even the creators of estimators usually do not prove that the supporting assumptions would make rational sense or common sense for the political actors being studied. [424, 436]

While Achen's critique has nothing direct to do with free will, his and others' desire to illuminate the importance of sound microfoundations does not stop when we write down

an *i* representing the individual, however. Whether or not humans possess free will, any social-scientific model without a solid, self-aware perspective on this issue is theoretically unanchored at the most "micro" level. Bracketing this debate would not generate any objections—if only political science models did not smuggle in tacit perspectives on this debate.

Brief philosophical background

Between stimulus and response, there is a space. In that space lies our freedom and our power to choose our response. In our response lies our growth and our happiness.

–Victor Frankl

The philosophical debate over free will, while likely familiar to most readers, is concisely illustrated and introduced by Table 1, which intersects views on whether human behavior is determined with the question of whether humans perceive that their behaviors are freely chosen:

Table 1. Nonpragmatic possibilities in the free will–determinism debate.

	Behavior is determined	Behavior is not determined
Humans make free choices	Compatibility (or "soft" determinacy) [3]	Liberty (free will) [1]
Humans do not make free choices	"Hard" determinacy [2]	"Hard"/double indeterminacy [4]

(Note that these are considered "nonpragmatic" possibilities, because a Pragmatist analysis of this debate may not even accept these dichotomies as meaningful.)

The four resulting views can now be summarized:

1. The (incompatibilist) libertarian, or voluntarist, view might be seen as the ordinary, "default" perspective most humans have of their behavior: their actions are not determined in advance by any outside force, but rather the result of "free" choices made by, or at least optionally made by, each individual in response to some degree of cognition.[5]

2. The opposing view, determinacy, argues that human behavior is entirely predetermined, and that our choices are not free. Hence, free will is entirely an illusion. This is sometimes referred to as "hard" determinacy, and shares an important assumption with the libertarian view: viz., that free will and predetermined human action are incompatible notions.

3. Compatibility, then, refers to a rejection of the idea that free will and predetermined action are incompatible, instead insisting that human actions can be predetermined even while humans retain something meaningfully called "free

[5] By "optionally made by," I mean that we might leave open the possibility that even while possessing free will, humans may, to one degree or another, (freely) choose to not reflect deeply on their choices, i.e., allowing subconscious forces, biological imperatives, etc., to guide their actions without introspection. Cf. assumption nine in section 1.4 below.

will," usually defined as "the ability to do what one wants" without problematizing how those wants are determined. This view is historically associated with Thomas Hobbes and David Hume, and more recently with Daniel Dennett.

4. Finally, the "hard" indeterminacy view represents the atypical position that humans lack free will, yet that our behavior is not fully determined, either. In other words, this view is based on acceptance of indeterminacy—such as quantum indeterminacy—that precludes predetermination of human action, even while denying that humans have any free control over this indeterminacy.

On a more erudite level, consider Inwagen's (1983) definitions of determinism and (incompatibilist) free will. He defines the former as "the thesis that the past *determines* a unique future … it is the thesis that there is at any instant exactly one physically possible future" (2–3). The latter is defined at greater length, but the essence is, first, that for an individual who has free will, "when he has to choose between two or more mutually incompatible courses of action … each of these courses of action is such that he can, or is able to, or has it within his power to carry it out …'[f]ree will', then, is to be defined in terms of 'can'" (8); and, second, closely aligning with my use, is the idea of "*immanent* or *agent* causation … [wherein] a certain change occurs in an agent, [who is himself] the cause of this change, and no earlier state of affairs necessitated this change" (4; all emphases in original).

Of course, this cursory overview covers little of the millennia of philosophical discussion of the subject; indeed, because of the extent of this debate in philosophy, this paper will intentionally not go beyond an "educated layman's" level of analysis of the debate.

One point of emphasis, however, is that there is no "middle ground" presented above. Each of the two dimensions of interest offer Boolean variables, forcing a clear and distinct response to each of the two questions. This fact is especially problematic for political science as is commonly practiced, given that research programs may either try to root themselves in such a muddled (were it even to exist) middle ground, or else weave back and forth by implication among the various distinct perspectives.

Autonomy

Another, perhaps more flexible, approach to thinking about free will is the concept of "autonomy" in human decision-making. Autonomy, though varying somewhat in definition based on different lines of research, in general refers to the ability of the individual to "govern" his or her behaviors in accordance with ideals and in spite of what can be seen as external desires.[6] Feinberg (1986) discusses four very similar meanings of autonomy: the "capacity" to govern oneself, the "actual condition" of self-government, an ideal of virtue derived from that conception, or the "sovereign authority" to govern oneself.

[6] For further insight into various ways of defining and relating the concepts of free will and autonomy, see Seebaß (2013) and especially Merker (2013).

The discussion of autonomy will dovetail, farther below, with the modeling technique of metapreferences, and on this note the idea of autonomy resembles, in some sense, the idea of objective internal deliberation—an internal voice, the superego, or a homunculus in one's head—that enables one to reflect, somewhat independent of biological and other material incentives, on behavioral options and make an autonomous choice. Even if copious material, societal, and other imperatives exist, what are the consequences of even a small degree of autonomy for political science research?

Causality

Causality is a complicated topic, but because the debate over free will frequently abuts into discussions of causality, I must at least briefly discuss by two metaphors this paper's perspective of how free will and causality relate.

1. Suppose a billiard-ball model of causality. The surface of the billiards table and the balls on it constitute the whole of the natural world, and the balls have been set in motion (perhaps perpetually, perhaps not) by some exogenous process. (The balls might be conveniently imagined to be some fundamental monadic components of the natural world, but this need not be the case.) All causality on this table is as simple, fundamentally, as our superficial yet straightforward perception and apprehension of the billiard balls striking one another and the table's edges, caroming off according to the simple laws of geometry and Newtonian physics, striking one another again, and so forth. An external observer or even one intelligent billiard ball might be able to learn about this natural world, understand material/efficient causality, derive natural laws, and so forth.

This is our basic model of causality: balls hitting walls and balls, on and on and on until (possibly) everything comes to a rest.

If we are but billiard balls, then social-scientific causality is no different from natural-scientific causality, and the way we interact with the world is predetermined. In this case, we might possibly understand the world, but we are powerless to alter its workings, and social and natural sciences stem from the same root.[7] (More on this below.)

On the other hand, suppose we are not actually balls but rather players with cues. Certainly, our shots will be influenced by the position of the balls (i.e., existing aspects of the natural world). We may have developed rules to determine which players can take what shots when, and that may preclude certain shots from ever occurring. Sometimes our cues might even be struck by the cues of other players—i.e., our freedom of behavior further constrained. And with our cues we cannot alter the shape or course of the billiards table walls at all; trying to do so would only result in broken cues.

Nevertheless, as players, we appear to stand outside the causal world of the billiard balls. Clearly, in this metaphor, what causes the billiard balls to move (including the hits made by the cues) is different from what causes the players to strike the balls. While there is an interface between—on the one hand—the natural world of the billiard balls, the table, and the cues and—on the other hand—the seemingly different world of the billiards players, the movement of the balls cannot be said to *determine* the actions of the players, even though the former may certainly influence the latter. Thus, in this model

[7] Or, put slightly differently, our power to alter its workings is entirely illusory; our alterations are, in fact, predetermined.

of free causation, there is a world (the table) that behaves completely deterministically, and there are players, some part of which lie outside that world, that can freely manipulate—i.e., interfere with—that deterministic world. While those actions are initiated outside the deterministic world, the deterministic world always maintains its law-abiding behavior, since the cues themselves (but not the holders) are part of the material world.

At this point, we might ask the following question: could we not observe the actions of the players over time with respect to the position of the balls, and determine a set of laws that the players obey—laws akin to those the balls follow? That is, after all, the basic proposition stemming from Hobbes and driving much of modern social science. Here we come to the essential move: suppose the billiards *players* are but billiard balls of a sort living in their own deterministic world, driven only by some second-order "social physics" equivalent to (perhaps ultimately identical to) the Newtonian physics at work on the original billiards table.

A social scientist, then, studies this second-order billiards table similarly to the natural scientist studying the original table. But again we might ask the question: are there not players at this second billiards table, standing outside the table yet interfering with it? Perhaps many of our social behaviors proceed deterministically, but don't we have some autonomous control of, e.g., how we might vote when on the fence in an election? Therefore we might imagine a second set of players influencing the second billiards table; yet if those players' actions are free, it implies that the players at the first

table are *not* like perfect billiard balls, because the second-order billiards game determining the players' actions at the first-order billiards game is not a closed, predetermined system.

So we presume to explain the behavior of players at the second billiards table via a third set of moving billiard balls on a third table. And, of course, this can go on and on, tables and tables and tables, with the behavior of the players at any given table explained deterministically with reference to the laws governing the balls at some further table.[8] The unavoidable result of this table-upon-table model of causality, given the working definition of free will above, is that there can be no free will; if we were omniscient, we could see that all billiard players are but billiard balls at another table, and in the end there is a table without any players whatsoever. With this, all the players throughout the system disappear, revealing a completely deterministic system that could be represented by one supreme billiards table—perhaps a sort of perpetual motion machine and Rube Goldberg machine come together. An intelligent entity within the system might be able to calculate the start- or end-state of the system or anything in between (as Laplace thought), but would be powerless to alter its workings.

The only alternative is that at some billiards table stand genuine players with a freedom of motion, decision-making, and so forth not circumscribed by any law-abiding

[8] Although this point will be emphasized later, I include "probabilistic" explanation as a subset of deterministic explanation, even though the two are sometimes conceived separately. In this sense, there may be occasions when an action kicks off a stochastic process; for example, in the context of this analogy, something may cause a player to randomly choose which ball to hit in accordance with some predetermined probability distribution function. However, generally speaking, macroscopic random processes are only random in the sense that humans cannot (due to lack of theoretical or practical knowledge or lack of computational ability) know the specific outcome in advance. On the quantum level, "true" randomness offers plausible evidence that even if humans lack free will, the physical world is indeterminate—i.e., the "hard" indeterminacy view noted above.

billiard-ball-like system. As I will elaborate below, this is a troubling conclusion for social scientific inquiry: it implies that past some point, our search for behavioral laws is confounded by the inscrutable inner workings of free humans.[9]

2. Second, suppose a diner poring over a long menu at a restaurant, famished yet indecisive. She wants to make a decision, but the menu is filled with several delicious-sounding options. Thankfully, a knowledgeable waiter stands by to help. The diner begins to ask questions about the dishes that look most enticing, and she quickly amasses some basic nutritional information on the various meals, forms improved expectations about what each meal will taste like, and takes careful note of the cost of each meal. All the while, she is evaluating her tastes: well, what am I hungry for?

As the waiter leaves to let the diner make a decision, the diner homes in on four dishes. Assume each meal is identical to all the others except for the following distinctions: meal *A* is expected to be exceptionally tasty, meal *B* is slightly more nutritious, meal *C* is made with ingredients grown in more environmentally friendly conditions, and meal *D* is slightly less expensive.

The question is, once the diner orders a meal, to what extent can we say that the attributes of the meal she chose—along with her specific situation—*caused* her to select a meal? More specifically, here, the question is how much "cause" is merely a vernacular

[9] A possible implication of free will, as illustrated in this metaphor for free causality, is some form of dualism. If we have free will, the "true" billiard players must be different in an important way from all the others: they must lie outside any manner of deterministic causation, in a somehow inscrutable portion of the natural world. There is also a nondualistic possibility: if all our attempts at understanding causation and determination have failed, the natural world could remain unified under indeterminism; however, rejecting any form of determinacy or comprehensible causality seems a taller order than adopting dualism.

term, and how much a meal's attributes "causing" the diner's decision is identical to a meal's attributes *determining* her decision.

In other words, if we opened the diner's head and could see sensations, perceptions, emotions, thoughts, and even individual atoms, before the diner placed the order, and if we knew all the diner's preferences, all the information the waiter would convey, etc., could we know what the diner would order ahead of time, and (equally importantly), could we say those sensations, perceptions, emotions, thoughts, etc., *completely determined* the choice? Or was there even an iota of uncaused, indeterminate, autonomous decision-making capacity in the diner's mind?

Plan of this paper

This paper is organized as follows. In the remainder of this introductory section, I will make explicit a number of assumptions I am making, then preemptively defend my main argument against several salient critiques.

In section two, I take aim at social-scientific arguments in the field of international relations that trace individual behavior back to either social structure (looking specifically at neorealism and social constructivism) or "thick" and "thin" models of rationality. I argue that most international relations research, by virtue of falling into one of these categories, must be rethought in light of the question of free will. I then consider a number of common dualities from IR research and discuss how the free will / determinacy debate better conceptualizes many of them. Finally, I point out some implications for international relations research if indeed humans lack free will.

In section three, I discuss the concept of metapreferences, especially in light of the problems of temptation and addiction. I then use these concepts as a launch pad to emphasize the role that science and technology play in human behavior and, indeed, social science if humans do have free will. I map out a few thoughts concerning how political science might be differently practiced if political scientists accepted the existence of human free will.

In the fourth section, I use the study of war—specifically analyzing Fearon's (1995) "rationalist" research—as a subject to be disassembled and rebuilt in light of the preceding discussions. I then close briefly.

Assumptions, biases, and related notes

As with any argument, the one made by this paper is supported by a network of assumptions. Listed below for transparency's sake are many of them:

1. Although the arguments in this paper apply generally to most, if not all, social science, the context for much of this paper is the study of international relations (IR). This presumes that IR theory is at least somewhat representative of social theory at large, and that the validity of the critiques of IR herein is not constrained to the present.
2. As noted above, this paper treats "compatibilism," i.e., the perspective that humans can possess free will even while all human behaviors are externally determined, as nothing more than a reframing of the meaning of "free will" in such a manner as to not comport with Inwagen's definition of the term adopted

above. Thus, this paper concludes that there is no way to reconcile free will in the indeterminate sense with determinism in human behavior, and treats the two concepts as entirely opposed and dichotomous.[10]

3. Pragmatic viewpoints on free will, such as those of William James (1896), are also ignored and treated outside the scope of this paper. Relevant or not, this perspective does not fit cleanly within the table presented in Table 1.

4. Likewise, this paper ignores hard indeterminacy as a social-scientific explanation, not because of its definitional incompatibility (pardon the pun) with free will and determinism as used herein, but rather as a simplifying assumption. Indeed, the presumably radical consequences of this form of indeterminacy have scarcely been considered in social science; for an exception see Wendt (2006). (On a somewhat related note, "indeterminacy" of the sort traceable to what may appear to us as randomness, but is actually due to the limitations of human knowledge / cognitive capacity [see fn. 8]—i.e., as is at the center of Gartzke's (1999) argument—is also not considered separately by this paper.)

5. The thesis of this paper is primarily to raise the relevance of the free will / determinacy *debate* for political science, not to come down strongly on a particular side of that debate. However, for various reasons, this paper generally adopts as a secondary goal emphasizing the consequences for international relations theorists of a general acceptance of the reality of free will, largely because in ignoring the debate, political scientists seemed to have erred more on

[10] For an interesting recent perspective in the philosophical literature, see Balaguer (2010).

the side of determinism. The second half (or so) of this paper therefore proceeds from (without an up-front defense) the opposite view.

6. It is easy to slip into the mentality that any instance of "selfless" behavior—for example, whenever one acts in contrast with his or her perceived biological instincts, is evidence for (or an overlapping idea with) free will. This paper makes no assertion that the two are specifically related, although cases in which humans act in contrast to biological imperatives *do* raise questions about the nature of human decision-making and how human behavior is determined. In this paper, to the degree possible, the emphasis is not on the nature of human behaviors, but whether external forces determine them. However, examples given herein are frequently situations in which one of two decisions a human faces is associated with fewer material/biological incentives than the other.

7. Perhaps most importantly, this paper will strive not to overanalyze what free will actually *is* (apart from the definitions provided in this paper) or how humans might ever have come to possess it. In many ways, thinking of free will is like trying to grasp sand tightly: it slips through ones fingers the more effort is exerted. This seems to be because in attempting to situate free human behavior in any sort of temporal sequence of events, it becomes cognitively very easy to lapse into thinking about free will as one step in a determinate causal chain. In a way, free will is the absence of something: the absence of anything connecting one physical event with another.

8. Related, this paper will make no strong commitments as to the questions such as, e.g., whether the "degree of freedom" of will (if it exists at all) varies from human to human as based on age, intelligence, or other factors. What is assumed is that if free will exists, then all humans must have some capacity for it. Closely related, and in accordance with Frankfurt (1971), I assume that no other natural entities have free will.

9. Finally and importantly, this paper does *not* assume that possessing free will would mean that humans cannot still undertake subconscious, habitual, reflexive, and otherwise unfree/predetermined actions. In fact, parts of this paper are based on the idea that in nearly any scenario, an individual human *may* have a materially/biologically/rationally/etc.-given "default" behavior, and that individuals can essentially suspend free will at times. Perhaps over the course of time a human can even "strengthen" or weaken his or her ability to make free choices, although this raises even further questions about free will.

Advance response to critiques

Finally, it may be useful to preemptively defend this paper's core argument, or at least its relevance, against some likely critiques. These abbreviated defenses are made in passing, and consequently should not be construed as exhaustive rebuttals.

Why so much stability?

First and foremost is sure to be a reiteration of Tullock's (1981) question, "Why so much stability?" In other words, if humans are free to think and act as they like, free of social-structural or material determinants, why do sociopolitical occurrences and outcomes appear, overall, essentially the same not only day-to-day but even across historical eras? Why do the same states persist across decades or more; why do leaders seek similar goals in conducting foreign affairs; and why do alliances, the breakdown of peace, and features of international trade exhibit similarities whether we examine the sixteenth century or the twentieth? In short, why does human behavior each day bear any resemblance to what it looked like the day before?[11]

An attractive, but ultimately not compelling, answer would be that free will does not necessitate that humans choose *different* behaviors over time, merely that the behaviors they choose are chosen freely. In other words, most individuals could be choosing to maintain the status quo, even though in principle these individuals could (in accordance with Inwagen's definition of free will) choose another path. While valid, the problem with a response such as this is that it is merely begging the question, or kicking the can down the way; in accordance with Occam's Razor, one might rightfully ask, "If humans freely choose their behaviors, but those behaviors are identical to what deterministic models would predict, what is the value-add of political scientists concerning themselves with free will?" Therefore a more sophisticated response—or,

[11] Ironically, instability can offer a perhaps clearly argument against free will. Kurzman (2004) argues that, in certain cases, "structure operates outside of conscious experience [and therefore] the subjective experience of unpredictability is discounted in favor of structural explanation—paradoxically … it is the structural account that restores agency to the [agents], whose subjective [anecdotal] accounts deny it" (346).

specifically, three responses—to this question is necessary. (Some of these responses foreshadow the later direction of this paper.)

1. The existence of free will would not require humans to (always) choose implausible (from a rationality standpoint) options. For example, consider again the diner model of causality given above, in which the point was made that because each choice was supported by some rationale, one of those rationales may be said to *cause* the ultimate choice. Therefore, on observing the diner over a long period of time, we may witness a type of stability: viz., a consistent range of finite outcomes, each of which accords with some preexisting rationale, and affords us a sense of long-term stability in the sense that whatever the exact pattern of outcomes, they all fit within an explanatory (perhaps even tautological) framework. That, however, does not imply or necessitate that *any* of the outcomes were predetermined by their "causes"; i.e., that the human element was unfree. For example, over a period of history we may be able to explain why some states formed in a certain manner and others in another manner, tracing these to a consistent range of finite outcomes, and through the historical lens "explaining" why, in each case, the outcome was consistent with the cause. This does create stability, but not (necessarily) stability in the sense that in each case the outcome is predetermined. Moreover, we may ignore events that did not happen, but would have (had they happened) been explained as though they were predetermined. Additionally, the way different decisions and human behaviors are defined can augment the appearance of stability. For example, when it comes to choices that

involve binaries, such as (simply defined) to go or not to go to war, it may lead create the appearance of stability in the sense that the history of humankind would show cycles between war and peace and perhaps consistent steps in the breakout of war and in the building of peace—even if none of those are predetermined. Together, all of this may lead to a form of explanation by "epicycles," along the lines of the model used in the ancient world to predict the position of the planets even while assuming an (essentially) geocentric system. Political scientists, with an arsenal of plausible "causes," can often explain most of the variance in a past or present event and relate a specific event to a general explanation. None of this appearance of stability, however, requires the behavior of individual human actors in the historical moment to be predetermined. Consider, for example, Waltz (121): "If the good or bad motives of states result in their maintaining balances or disrupting them, then the notion of a balance of power becomes merely a framework organizing one's account of what happened, and that is indeed its customary use. A construction that starts out to be a theory ends up as a set of categories. Categories then multiply rapidly to cover events that the embryo theory had not contemplated. The quest for explanatory power turns into a search for descriptive adequacy."

2. Freely willed behaviors would still operate amid a backdrop of unyielding material constraints (a claim which will be very relevant in section three below). In other words, even if humans possess free will, no human can (at this point in time) will herself to be a carrot, or will himself to be immortal. Even if the

objective world of physical laws allows some form of free will, most libertarians would not argue that this nullifies any determining physical laws, such as the physical laws governing most aspects of human biology, the physical geography and geology of the earth, etc. Therefore we may frequently see forms of stability that are simply traceable back to these underlying physical realities. For example, politics may frequently concern issues of land, food, the environment, health, etc., simply because these are the objective physical realities we share. Note that this does not imply that human behavior toward these physical realities must (in a predetermining sense) take a particular form, such as fighting to acquire more and more of scarce resources. Additionally, human knowledge may be considered to be under the constraints of the natural world; we cannot will ourselves to omniscience, and therefore the growth (or decline) of human knowledge may generate (or obstruct) political concerns on a somewhat predictable timeline.

3. Freely willed behaviors can be "nested," with some falling in accordance with some socially given constraints. Consider Tullock's broad explanation of stability: "In some cases … stability will not be a true equilibrium because a random member of a large set will be chosen and then that random outcome will be left unchanged for long periods of time" (189). Thus, stability may also be derived from certain elements of the human experience simply being on "autopilot" for periods of time.

4. Theories, models, and even apparent "facts," may, somewhat arbitrarily, create an illusion of stability where none exists. For example, Waltz (1979)

discusses the pervasiveness of the operation of the balance-of-power in international (read: interstate) history. However, in the course of world history, contexts in which the balance of power did not operate would likely have seen consolidation among political entities, hence "hiding" them from the criterion of "international" used to scope where balances of power are supposedly forming.

5. Perhaps most importantly, the possible existence of free will would not prevent humans from still having rational goals and thereby acting out behaviors that bear logical connection to those goals, and that are logical in coordination with one another. In other words, we may still expect enmity between states to be a necessary (but not sufficient, not predetermining) condition for war, and by holding enough constant, we may be able to predict with reasonable success some outcome. If what we hold constant are human decisions/actions subject to free will, however, then the likely outcome is not necessarily likely given changes in what is being held constant. Nevertheless, we may be able to draw a conclusion along the lines of, "97% of likely voters prefer Lincoln to Douglas; therefore we predict Lincoln will be elected."

"Prediction is all that matters" or "Empirical accuracy is all that matters"

What about the idea that a theory or model need not be well-grounded philosophically or have its inner workings resemble reality in order for it to be empirically accurate and, therefore, relevant; viz., the argument laid out famously by Friedman (1953)?

This paper takes the view of Cohen (2000), who in defending analytical Marxism, notes that

> a micro-analysis is always desirable and always in principle possible, even if it is not always possible to achieve one in practice at a given state of the development of a particular discipline. … [P]re-analytical Marxism was scientifically undeveloped, rather in the way that thermodynamics was before it was supplemented by statistical mechanics, and, in each case, because of failure to represent molar level entities (such as quantities of gas, or economic structures) as arrangements of their more fundamental constituents. It is one thing to know, as phenomenological thermodynamics did, that the gas laws hold true. It is another to know how and why they do, and that further knowledge requires analysis … . [xxiii–xxiv]

Similarly (and as quoted by Wendt [1997, xvi]), Whitehead notes, "No science can be more secure than the unconscious metaphysics which tacitly it presupposes."

Thus, even if political scientists had perfectly predicting models (and, clearly, we do not), a microfoundational understanding of individual actors' behavior can contribute to our understanding of how and why political-scientific phenomena behave as they do.

Additionally, and as will be discussed in more detail further, social scientists must recognize the centrality that the volition debate has for normative concerns, such as what responsibility social scientists have to promote positive change in social systems, or even whether we have a choice in the matter of what happens in society.

"Free will is just an error term"

Another potential critique would posit that even if free will exists, its effects would be similar to an "error term," i.e., representing stochastic deviations from an otherwise stable and/or predictable mean. There are several defenses against this line of critique:

- It assumes we know something about the distribution of such errors: presumably that they have a conditional mean of zero, i.e., that the errors somehow cancel each other out. This might make some sense within the context of a purely quantitative mode, but what about in qualitative context? Even if freely chosen human actions are merely "random" deviations from what is predetermined, these actions may often not take the form of one-dimensional numeric errors. These errors may be, rather, something like the behavior of the "Tank Man" who attempted to block the column of tanks entering Tiananmen Square in 1989. Even if simple "errors about a mean," the cascading consequences of such errors may shift the mean itself, and it is unclear why we should assume these errors would cancel out across many dimensions.
- It assumes the "mean" is not only knowable, but can be distinguished from the errors in empirical analysis.
- Finally, it raises questions about the nature of free will: how can human decisions determinately arrive at a consistent, mean outcome, yet be free to contain occasional errors? In other words, how and why could some decisions be predetermined and others free?

"We should study what we know" or "This is for philosophers to sort out" or "Even philosophers don't understand free will"

Given the preponderance of philosophical argumentation in political science, and especially the study of international relations, this critique seems unlikely to be persuasive from the start. However in brief rebuttal, I would simply point out that even if it is up to philosophers to tell us what does and doesn't make sense in the context of the free will debate, political science theories and models should be expected to comport with *some* plausible viewpoint on most, if not all philosophical issues. Again, the argument in this paper is that political science models are muddled on the debate, coming down unclearly and inconsistently on different sides without showing a genuine comprehension of either the state of research on the topic in the philosophical literature or the implications of that research for political science.

Additionally, we may find viewpoints and arguments concerning free will and determinism overly abstract and unverifiable. Why should such seemingly far-from-the-here-and-now topics concern political scientists? This paper takes the view that the free will debate is similar to debates over epistemology, the nature of causality, and the like, all of which go far deeper than the day-to-day focus of many political science models. Nevertheless, these ideas must explicitly or implicitly undergird all such models, and therefore all models can be critiqued on this front. For example, there is a good deal of talk in international relations theory about the debated connection between norms and state behavior. But how does a norm actually influence any individual's behavior? How does an idea, both in subjective cognition and in objective brain chemistry, actually lead

to the outcome of behavior? Such questions lie just beneath the surface of essentially every social science theory.

Finally, even if political scientists were adopting standard assumptions so as to validly "bracket" the free will debate and search for theoretical conclusions that would stand no matter the result of that debate in philosophy, we might expect political scientists to state or defend such assumptions transparently at least once in a while, even if only in pedagogical materials. Instead we find relatively few mentions, let alone deep analyses, of the topic at all.

"The idea of free will leads to a (fallacious) infinite regress"

Finally, some may argue that any discussion of free will is essentially a nonstarter from the logical standpoint. Ryle (1949), for example, makes a point in his critique of intellectualism: that to say that any intelligent behavior requires, first, a thought regulating that behavior, merely leads to a regress in necessitating a second thought to explain the origin of the first thought, and so on.

It should suffice to say that other philosophers have critiqued this viewpoint (e.g., Stanley and Williamson 2001), and therefore this may be taken as showing enough philosophical breathing room remains for theories that accept free will. This breathing room may exist in conjunction with other specific viewpoints, such as Cartesian dualism, and thereby accepting free will may require or entail accepting other potentially controversial, but plausible, philosophical points of view.

Finally, it bears keeping in mind that this paper does is open to the idea that humans lack free will and that our behavior is predetermined; but even if so (e.g., even if nearly any of the critiques above are valid), political scientists must reconcile their theories with fundamental determinism.

Chapter 2: The Problem of Free Will for IR Theories and Concepts

In order to explain the importance of considering the free will debate in international relations theory, this paper must first explain how the definitive existence or nonexistence of free will would eviscerate the most popular IR paradigms of much of their theoretical relevance. Focusing on certain exemplar texts, the following two subsections hold, first, structural and other holistic IR theories and, second, individual-level rational-choice and psychological theories, up to the light of the free will–determinism debate, arguing that all of these theories do not well cohere when seen from this perspective. Farther below, this paper argues that the free will–determinism spectrum subsumes and more transparently represents several dualities commonly used in IR.

"The system made me do it": freeing the individual from society/structure

If all social behavior occurs, ultimately, only through the actions of individual humans, then to understand social outcomes we must think carefully about what, if anything, determines an individual's actions. In this section I will both elaborate on and defend this premise, more fully expanded as follows.

1. Social structures only exist in agent's heads; their physical manifestations are produced by individual agents; structures can only be created, reproduced,

altered, etc., through some form of agent action. As Adler and Pouliot (2011) neatly put it, "[S]ocial structure does not cohere on its own" (16).

2. It is primarily useful, from a social-scientific standpoint, to think of social structures insofar as those structures are somehow determining the behavior of agents operating within it. (One may argue that it is also useful if speaking of structures simplifies the process of explanation without losing accuracy; however this paper is challenging that for this to be true, structures *must* be determining unit behavior.)

3. If humans possess free will, then this not only nullifies any social structures' determination of human action, but also calls into question the degree of influence such structures have over human behavior. Explanations that focus on the group-level will not be accurate except insofar as they are heavily constrained with specific context and assumptions; even then,

4. If humans lack free will, then the physical processes by which causality is transmitted call into question what relevant role social structures and explanation based on them have; instead, these structures are fully epiphenomenal, mere subjective representations that add nothing to the explanation of human processes.

This line of thinking applies to both social structures and societies; that is, whatever definition or idea of a multiperson (or multi-entity) grouping is used. In either event, research tends to reify social structures and society at large, substituting them for social networks (i.e., agent-to-agent relations). Set theory gives a good illustration of this criticism. Thinking of society casually as a monolith is like thinking of a single set *S* with

N individuals as members. Thinking of society as layers of networks and relations is like thinking of the power set of *S*—i.e., the set of all possible sets that can be constructed from the members of *S,* viz., much greater than *N* (and increasingly so as *N* grows). Society is not the simple monolith, but rather networks, and networks of networks, few if any of which contain all members of society, and each with their own social structure. Whenever "society," "system," and "structure" appear in political science, what we actually have is shorthand for some collection of agent-to-agent relations. The objection is not to the use of this "shorthand" *per se*, but rather to the reification this frequently facilitates.[12]

This is not to deny structural/systemic/social viewpoints a voice straight out of the gate. As Hafner-Burton et al. (2009) argue, social networks can still "define, enable, and constrain those agents" whose relations constitute such networks. But just as the state system is, in a sense, a figment—albeit one genuinely encountered, along the lines of Wendt's (2010) hologram—societies and networks too may genuinely "define, enable, and constrain," but not without saying something, and perhaps a great deal, about the agents themselves.[13]

This is the essence of the debate between Althusser and Thomspon, and of another between Lévi-Strauss and Sartre; it also sits at a crucial point in Giddens (1984): are agents merely the passive vessels of social structures, or do they have any autonomy

[12] Cf. Berger and Pullberg (1965).
[13] These comments apply as well to Adler and Pouliot (2011) and Jackson and Nexon (1999) also. Although this paper does not specifically respond to relatively recent theoretical developments that emphasize practice/process over more traditional objects like structures, they are nevertheless susceptible to many of the critiques in this paper, in that they raise the question: do practice/processes determine social outcomes, or are humans always free to resist their influence?

to respond to social structures in their own way? And if they have any autonomy, then in what ways or senses do they not have complete autonomy?

First, if agents are completely passive, as in a good deal of Marxian and structuralist thinking, then there is a biconditionality between structuralism and determinism: the accuracy of structuralism implies agents are predetermined by the structure to act in a particular way (even if counterstructural), while the reality of determinism implies humans lack any agency to take any meaningful stance with respect to structure at all. Yet there is something of a paradox here: if structures indeed determine agents' actions, it can only be because they agents are fundamentally unfree to act any other way. Thus, structuralist theories (e.g., Waltz [1979]) are saying quite a lot about the unit level, because all the determining work done by structural theory must be mimicked by tacit determinism on the unit-level. As Wendt (1987) describes, "[I]n both its decision- and game-theoretic versions neorealism, like microeconomics, is characterized by "situational determinism," by a model of action in which rational behavior is conditioned or even determined by the structure of choice situations." (342).[14]

Second, what if agents are not completely passive, and have even the slightest bit of freedom within the constraints of structure? Here, the obvious question is how agents can not be just a passive bearer of structure, yet be in *any* way controlled by it. That is, what microfoundational model of man gives us *genuine* freedom to act within a certain context but denies it to us in other contexts? Often, these views seem to suggest the equivalent of saying we have "free will every Tuesday": we have either some measure of

[14] Wendt also insists that Waltzian neorealism is more reductionist than Waltz claims.

free will, or else complete free will but only in some times or some situations, and when it is convenient for the theoretical approach, the freedom of human will magically disappears. How can we be *partially* free?

In this sense, structurationism and process theories may improve on structuralism by reconciling the agent–structure debate more closely with our personal experience, yet they muddy the waters to some degree. Saying that agents and structures are "mutually constitutive," for example, is preferable to saying that structures do all the constituting and causing (i.e., reifying a monolithic society that inhumanly possesses agency while denying it to actual humans). But by what mechanism can social networks, practices and processes, and other agent-to-agent relations determine an agent's actions at all?

Consider, for example, these passages from Hopf (1998), Jackson and Nexon (1999), and Adler and Pouliot (2011), respectively (especially where I have emphasized):

> How much do structures *constrain and enable* the actions of actors, and how much can actors *deviate from the constraints* of structure? … [I]dentities and interests are produced through social practices … An actor is *not even able to act as its identity* until the relevant community of meaning … acknowledges the legitimacy of that action, by that actor, in that social context. [172, 176, 178–9]

> [P]ractices are both individual (agential) and structural. When 'disaggregated', practices are ultimately performed by individual social beings and thus they clearly are what human agency is about. Collectively, however, we understand practices as structured and acted out by communities of practice, and by the diffusion of background knowledge across agents in these communities, which similarly *disposes them* to act in coordination. Practices are agential, however, not only because they are performed by individuals and communities of practice, but also because they

> frame actors, who, *thanks to this framing, know who they are and how to act in an adequate and socially recognizable way*. … Recursively, in and through practice, agents *lock in* structural meaning in time and space. Agency also means doing things for reasons, many of which are *structurally supplied*. … While performed by individual human beings, practices are possessions of collectives *insofar as their meanings belong to communities of practice*. [16]

> Agent–structure problems are concerned with the causal or constitutive relationship between individual actors and aggregate social forces. For instance, to what extent does social structure *constrain or enable* individual choice? To what extent is social structure autonomous from agents? Perhaps the most influential 'solution' to the agent–structure problem has been introduced into IR by constructivists … . In this formulation, agents and structures are 'mutually constitutive' … agents instantiate structures through their actions, even as those structures simultaneously *constrain and enable* agency. [295]

In each case, while the "compromise" approaches attempt to situate agent and structure so that they can co-exist harmoniously (at least for the academician), the structure still somehow seems to be robbing the agent of free will on weekdays and every third Saturday.

None of this should suggest that agents' actions cannot influence other agents' actions, or that one agents' actions may not alter the physical realities that delimit options for other agents. And these physical realities do, of course, intertwine with what may be called social realities. Social networks, process and practices, and other agent-to-agent relations clearly have something to do with individual choice at least some of the time. But the sort of influence is directly contrary to the thrust of most structural, structuration, constructivist, and even process/practice thinkers; the influence is, ultimately, under the

free individual's control to modify. For instance, entirely contrary with Hopf's declaration that an actor's identity is contingent on social acceptance are the existentialist and solipsistic perspectives: an actor's identities, decisions, etc., are mediated in an intimate, entirely personal, and perhaps *sui generis* process for each individual, and even assuming that any one individual's experience is genuinely *like* any other's is a major assumption.

This paper does not stray into subjectivism, however, because the free will argument does not require that each individual be completely autonomous in a socially ontological sense. Rather, I think the most important point is that free will implies that each individual has decisional autonomy, and—referring the diner model of causation from above—how social networks and the like "cause" individual behavior is far more nuanced than the above critiqued theories usually explore.

To help present the remainder of my argument for this section, I would like to break structure-aware theorization into two broad groupings. The first I will nickname "hard" socializing theories, epitomized and represented below by Waltz (1979); these "market-like" theories do not emphasize systemic/structural/societal construction or generation of agents/actors, merely that the holistic level of analysis suffices to explain a broader level of outcome. Second are "soft" socializing theories, along the lines of Wallerstein (1979) and especially Wendt (1999). (In fact, Waltz and Wallerstein are compared and critiqued by Wendt [1987], and my lumping-together of the latter two represents my view that structurationism does not solve the fundamental questions

underlying how individual agency relates to social structure.) Ultimately, I will argue that both forms of socializing theories converge on the same question of behavioral determination.

Hard-socializing theories: structural realism

First, consider hard-socializing theories. In these, structures/societies/practices/etc. do not actually constitute or modify actors' identities, preferences, etc.; rather, the dynamics of strategic interaction, supposed systemic imperatives, selection effects, and the like arguably determine individual behavior and thereby social behavior. Structure is therefore a context that, according to structural theorists, is inherent in unit-to-unit interaction rather than something that alters the units *per se*.

This section uses Waltz's seminal *Theory of International Politics* as the epitome of structuralist theory, and representative of the problems of this line of thinking in light of the free will / determinism debate. Before beginning, however, this paper must answer a question: how is the question of human will relevant given that Waltz, and other neorealists, rely on the unitary actor assumption and focus on *state*, rather than human, behavior? Without delving into a full defense, on this point this paper stands on the idea that the unitary actor assumption must itself be defended based on logic that is, ultimately, very similar if not identical to the logic of structural theories of international relations. (At times, this paper leans on the unitary actor assumption and endows "free will" to the collective of a state.)

My goal in this section is to show that Waltz is internally inconsistent when examined through the lens of the volition debate.[15] I will use this to counter Waltz's promotion of structure, as he claims, "Political scientists … reify their systems by reducing them to their interacting parts" (61).

1. Are states free and undetermined in what they seek?

The first question to be analyzed is how much freedom Waltz grants states. In some passages, Waltz acknowledges that states' fundamental interests may vary and indicates an acceptance of indeterminacy in the formation of state interests:

> Internationally … the structure of the system … is set by the fact that *some states* prefer survival over other ends …. [93]

> Theory explains regularities of behavior and leads one to expect that the outcomes produced by interacting units will fall within specified ranges. The behavior of states and of statesmen, however, *is indeterminate.* How can a theory of international politics, *which has to comprehend behavior that is indeterminate*, possibly be constructed? [68–9]

> [T]hough balance-of-power theory offers some predictions, *the predictions are indeterminate*. Because only a loosely defined and inconstant condition of balance is predicted, it is difficult to say that any given distribution of power falsifies the theory. … [A]lthough states may be disposed to react to international constraints and incentives in accordance with the theory's expectations, *the policies and actions of states are also shaped by their internal conditions*. [124]

Elsewhere, his language indicates that there are incentives, or pressures, for states

[15] In all excerpts here, emphasis added except as noted.

to be obsessed with survival. Notice that the language, especially near the end of the second passage, quickly transits from the assertion of worry on the part of synecdochical states to claiming that worry translates into behavioral changes:

> In an unorganized realm each unit's *incentive* is to put itself in a position to be able to take care of itself since no one else can be counted on to do so." [107]

> Insofar as a realm is formally organized, its units are free to specialize, to pursue their own interests without concern for developing the means of maintaining their identity and preserving their security in the presence of others. … *In any self-help system, units worry about their survival, and the worry conditions their behavior.* [104–5]

Eventually, Waltz claims that states are full-on restricted in their attitudes and actions:

> [N]ations … *must be* more concerned with relative strength than with absolute advantage. [106]

> [S]tates *have to do* whatever they think necessary for their own preservation, since no one can be relied on to do it for them. [109]

> A balance-of-power theory … begins with assumptions about states: They are unitary actors who, *at a minimum, seek their own preservation* and, at a maximum, drive for universal domination. [118]

Waltz likewise leans on the standard logic that the security dilemma is inescapable for states (186–187).

Therefore on a very fundamental level with respect to the volition debate, Waltz is unclear. To what degree, in his model, are states (and, by extension, their constituents) permitted to make autonomous decisions with respect to their behavior, versus to what

degree are they expected to, for undefined but assumed-to-be-predetermined reasons, fall lock-step into certain classical patterns?

2. Do states understand what they're doing?

Another crucial question is whether states are able to see themselves inside of a structure, and the degree to which they understand how their actions may perpetuate that structure. This is highly relevant to the question of free will, as the ability to fully (or even partially) understand the context and consequences of their actions as they effect the structural constraints that bind them and other states suggests that states have the opportunity to reflect meaningfully as to the relationship between their actions and the structural context in which they act.

For example, at one point Waltz is clear that states do not intentionally/knowingly abet the structure in their actions:

> Structures emerge from the coexistence of states. *No state intends to participate in the formation of a structure by which it and others will be constrained*. [91]

Elsewhere, the structure is seen as independent of actors' knowledge about structure:

> Insofar as selection rules, results can be predicted whether or not one knows the actors' intentions and *whether or not they understand structural constraints*. [76]

> Actors may perceive the structure … [but] may for any of many reasons fail to conform their actions to the patterns that are most often rewarded and least often punished. [92]

Yet in examples, such as the discussing the security dilemma (186–7) or in the analogy of the run on the bank (107–108), Waltz implies or clearly states that the actors know full

well the structural environment and the effects their actions may or will have, and conduct reasoned cost–benefit analysis before committing themselves to an action.

3. Is structure an agent?

What about whether structure should be thought of, however figuratively, as having independent agency? On this point Waltz begins clearly:

> Structures do not work their effects directly. Structures do not act as agents and agencies do. How then can structural forces be understood? How can one think of structural causes as being more than vague social propensities or ill-defined political tendencies? *Agents and agencies act; systems as wholes do not. But the actions of agents and agencies are affected by the system's structure. … Structure affects behavior within the system, but it does so indirectly.* The effects are produced in two ways: through socialization of the actors and through competition among them. [74]

However, he soon lapses into the language of structure and structural processes possessing agency of their own. (Granted, this may be linguistic informality, but it bespeaks adherence to notions that are more troubling, discussed below.)

> The first way in which *structures work their effects* is through a process of *socialization that limits and molds behavior*. The second way is through competition. … Socialization *encourages* similarities of attributes and of behavior. So does competition. Competition *generates* an order, the units of which adjust their relations through their autonomous decisions and acts. … Insofar as selection rules, results can be predicted whether or not one knows the actors' intentions and whether or not they understand structural constraints. [76]

> The problem … is to contrive a definition of structure *free of the attributes and the interactions of units.* Definitions of structure must leave aside, or abstract from, the characteristics of units, *their behavior*, and their interactions. [79]

> Structures *encourage* certain behaviors and *penalize* those who do not respond to the encouragement. [106]

Finally, Waltz is at times unclear about the degree to which structures should be said to be agents:

> *Structure may determine outcomes* aside from changes at the level of the units and aside from the disappearance of some of them and the emergence of others. … *If structure influences without determining, then one must ask how and to what extent the structure of a realm accounts for outcomes* and how and to what extent the units account for outcomes. [78]

Again, while this may be brushed aside as linguistic convenience, this paper argues that the inconsistency belies a muddled perspective on how independent structure is in terms of its ability to serve as, effectively, an agent in the international system. Even the casual use of language allows Waltz to set apart the effects of structure as somehow occurring independent of the actions of the states that comprise the structure; buried between the lines is something akin to an assumed bootstrapping process whereby states at time t are presumed to have acted in such a way to create a certain structure at time $t + 1$ which then *independently* constrains the states in that period, who then, by virtue of that constraint, have no choice but to perpetuate the same broad structural principles that persist to time $t + 2$ (with the same effects on the states in that period), and on and on. An example of this thinking is obvious when Waltz writes that

> Competitive … international-political systems work differently. Out of the interactions of their parts they develop structures that reward or punish behavior that conforms more or less nearly to what is required of one who wishes to succeed in the system. … *Patterns of behavior ... emerge, and they derive from the structural constraints of the system.* [92]

Clearly Waltz does not genuinely identify the international structure as an autonomous agent, able to act in the complete absence of its constituent pieces, any more than theorists would identify the state as able to act in the absence of its constituent humans. Nevertheless the language indicates a fundamental dodging of the central question that the volition debate brings into focus: if structure compels states into necessary actions, i.e., if it *determines* state action, then states (and, more importantly, their human constituents) must at some point be denied the ability to freely resist structure's effects. If, on the other hand, states and their constituents can freely resist structural incentives, the real theoretical value of the structural model is threatened.

4. Does structure stand on its own?

Pushing this point farther, consider the question of whether structure "stands on its own," or whether it is integrally tied to specific behavioral choices on the part of states.

For example, consider Waltz's double standard when it comes to discussing the possibility of international governance. At one point, he merely speaks in terms of the *incentives* that, arguably, influence state actions in the context of centralizing international power:

> The more powerful the clients and the more the power of each of them appears as a threat to the others, the greater the power lodged in the center must be. The greater the power of the center, the *stronger the incentive for states* to engage in a struggle to control it." [112]

Speaking of incentives, as elsewhere, is more understandable. Yet in introducing this discussion on the preceding page, Waltz asserts much more strongly—and in doing so, gives structure independent agency—that these incentives would prevent world government:

> In a society of states with little coherence, attempts at world government *would* founder on the inability of an emerging central authority to mobilize the resources needed to create and maintain the unity of the system by regulating and managing its parts. [111-112]

Another relevant example in which Waltz is inconsistent in attempting to transition from micro-behavior to structural constraints is in his analogy of the demise of the corner grocery. In the early stages of this analogy, Waltz notes that:

> If the market does not present the large question for decision, then individuals are doomed to making decisions that are sensible within their narrow contexts even though they know all the while that in making such decisions they are bringing about a result that most of them do not want. Either that or they organize to overcome some of the effects of the market by changing its structure …. [108]

Waltz says inconsistently that

> [I]ndividuals can do nothing to affect the outcomes. Increased patronage *would* do it, but not increased patronage by me and the few others I might persuade to follow my example. [108; emphasis original]

That is, the individuals are fundamentally able, by virtue of their decisions, to change the structural outcome, but Waltz immediately rejects the possibility that individual-level decisions can have a net effect. Waltz allows himself and a few others (in the analogy)

freedom of action, but an unseen hand ensures that their numbers will remain few enough that "individuals can do nothing to affect the outcomes. Hence, eventually, Waltz simply asserts that:

> The only remedy for a strong structural effect is a structural change. [110]

This, despite the fact that in his analogy there was clearly a phase during which the "problem" could be remedied by certain individual action; viz., each individual choosing to patronize the corner grocery instead of the centralized supermarket.

Lastly on this point, Waltz's analogy of a bank run, pp. 107–8, falls victim to the same critique. Waltz rushes from discussing the Prisoner's Dilemma-esque nature of each individual's situation on the eve of a bank run to insisting that the problem is *structural* and therefore not able to be overcome by virtue of how individuals play the dilemma.

The key problem, especially upon examining Waltz's examples on page 122 regarding the theory of the firm/state vs. the theory of the market/international system, is that Waltz claims to be simultaneously (1) ignoring how states function (i.e., come up with foreign policy, as he notes that his isn't a foreign policy theory; in his words, "The theory makes assumptions about the interests and motives of states, rather than explaining them"), yet he insists that (2) he is saying what "constrains" states —in his words, "What it does explain are the constraints that confine all states."

5. How robust is the system/structure?

The ultimate point concerns the durability, or robustness, of the system, as defined by Waltz. How will it handle in the face of varied individual behavior? The

remainder of this subsection will focus on this question, as it best represents the problem with hard-structural theories: tacitly positing the persistence of structure even while remaining fuzzy on how much freedom of action the individual units possess.

At one point, Waltz notes usefully:

> Obviously, the system won't work if all states lose interest in preserving themselves. It will, however, continue to work if some states do, while others do not …. [118]

And, as quoted above, Waltz indicates:

> Actors may perceive the structure … [but] may for any of many reasons fail to conform their actions to the patterns that are most often rewarded and least often punished. [92]

Yet what reveals the inconsistency is his analogy to the path of a falling leaf:

> True, [balance-of-power] theory does not tell us why state X made a certain move last Tuesday. To expect it to do so would be like expecting the theory of universal gravitation to explain the wayward path of a falling leaf. [121]

Here, Waltz implies that states' foreign policy behaviors—i.e., the equivalent to the forces affecting the falling leaf's "wayward path"—are unrelated to the (systemic) forces creating the general tendency for state's foreign policy—i.e., the universal constant of gravity. As reviewed above, this places a false wedge between what is, in reality, one and the same force.

My attack on Waltz's logic here is helped along by some who are at least sympathetic to this form of thinking. For instance, Feaver (in Feaver et al. 2000) acknowledges,

> If realists expect some states to flout realist principles—indeed, expect democratic states to be prone to do so—and if the number of those states grows exceedingly large, is it not possible that at some point most states are not behaving according to system constraints? If that happens, what is

> left of the system to enforce the constraint? Can a universe of system-ignoring democracies literally invent a novel set of system constraints? Constructivists have no problem answering in the affirmative, but realists surely are inclined to answer in the negative. Realists, after all, do argue that some state goals (though not all …) are irreducibly conflictual. [168–9]

While we must sort through the built-in unitary-actor assumption in this excerpt, the thread of determinism at this unit-level is unavoidable. Assume a world of *N* states, in which each state is offered a choice between supporting and defecting from some sort of "systemic" behavior. Realists acknowledge that states {1,2,3,…,i-1} may defect for various reasons. But for some reason, by the time we get to states{i,…,n-2,n-1,n}, these states—which may be undifferentiated from the ones that defected—are somehow unable to defect, either robbed of the choice or unable to make the choice freely. But clearly the phantom system cannot rob states of the choice, so by the time we get to state n, we have to assume decision-makers are actually, physically unable to defect. They are immune to whatever ideas, norms, irrationality, domestic factors, etc., allowed the earlier states to defect.[16]

What usually seems to be going on in such theories is exemplified by the logic of the sequential Prisoner's Dilemma, albeit with "defection" here relabeled to mean "not defecting," i.e., the cooperative-yet-irrational behavior in the Prisoner's Dilemma. As

[16] This point drives against the realist caricature of likening states ignoring realism's dictates to individuals jumping off a building and thinking they will fly. This clearly presumes some external actor (either the system or gravity) that is somehow different and could not "jump off the building," too. And the relevance of this criticism is also tied in with whether realism is a normatively good theory (because it helps leaders avoid disastrous attempts at harmony) or bad (because it scares leaders into acting shrewdly). Also in this note I just want to mention that a separate defense of realism is its alleged empirical usefulness, rather than its theoretical validity or success. I will deal with this sometime. Some structural realists give more attention to such issues. For example, Snyder (1997) explains, "[S]ystem structure "must pass through a domestic politics prism consisting of the perceptions and values of decision makers and the domestic constraints that bear on them. Systemic constraints make themselves felt only through people who make decisions, even if, sometimes, those people are not fully aware of them" (131).

states begin to cooperate in a hypothetical *N*-person Prisoner's Dilemma, the benefit of refusing to cooperate (i.e., obeying the systemic imperative) grows. Materially grounded thinkers, such as political realists, seem to suggest that the temptation to behave "realistically" grows to the point of irresistibility by the time the n^{th} state (or politician) has moved. But in suggesting the existence of such irresistible temptations, there is tacit determinism.

There are two reasons this is problematic. First is simply, as this paper argues throughout, that realists must lean on determinism—and invariably individual-level determinism—in order to make their explanations stick. But this reliance is often opaque and, were it not, would likely be controversial, both because it is deterministic in general (i.e., not compatible with belief in free will) and because of the specific content of the determinism (i.e., that people's behavior is determined in particular pro-structural ways). Second (discussed farther below) is the issue of whether, and if so, how, structural theories can remain coherent at all given the predictions they make and such tacit determinism. Plainly put, can this sort of structural theorizing explain why the system inevitably reproduces *and* explain why there are some defections, without being completely tautological?

The determinism must run down to the individual level because, of course, unitary actor frameworks simply bracket domestic processes, usually on some assumed grounds. For example, unitary-actor realists might argue that domestic selection effects invariably force states to behave as the system "wants" them to. This may mean that when it comes to Prisoner's Dilemma-type situations, a state can genuinely be thought of as losing

substantial utility for playing the "sucker" because the domestic leader will be kicked out of office. But within the domestic system, we face the same sort of questions: why must such things be inevitable? (Of course, such arguments are not original to this paper; students of individual psychological and bureaucratic politics, and analytical liberals, especially, have focused on the bottom-up process of preference aggregation. For those types, this section of my paper is somewhat irrelevant anyway, since they do not privilege structural or systemic explanations of social behavior.) Inevitably, and in opposition to, e.g., Waltz (1979), for any particular systemic framework to be useful, we must assume a good deal about human nature and its constancy, in the sense that our nature *determines* some actions beyond our control. Take, for example, one of Grieco's (1988) arguments against neoliberal institutionalism, the issue of relative gains. Grieco asserts,

> [The] realist-specified function does not suggest that any payoff achieved by a partner detracts from the state's utility. Rather, *only gaps in payoffs to the advantage of a partner do so*. The coefficient for a state's sensitivity to gaps in payoffs—k—will vary, but it will always be greater than zero. [501; emphasis original]

Obviously, our experience suggests that people often exhibit such "raw selfishness," but what grounds does Grieco have for forbidding states from *ever* being charitable in gains. Even if we have never observed unquestionable state charity empirically (I'm sure there would be some lively debate over such a question), our experience as individuals seems to suggest that, at least in theory, k could vary between –1 and 1 instead of being strictly positive, as Grieco argues. Yet Grieco's claim fits well with realists' general assertion of

a sort of "no true selflessness" theorem.[17] An alternative conception is the belief that people do follow ideas, but only opportunistically; but as with "no true selflessness," the problem is that the "opportunities" under which people follow ideas usually expands to situations that conceivably threaten the supposed iron systemic imperatives.

One last illustration of this point is the famous "theater-goers fleeing from a theater fire" example. While there are many critiques of this argument, such as in Hopf (1998), they stand apart from the line of critique offered herein. Viz., the issue is not just whether there are one or multiple exits, whether there are issues of "who goes first," etc., but whether a theater fire actually *causes* the flight in a closed, deterministic sense. The experience of Bouazizi and Duc, discussed above, suggests that human reaction to fire is mediated by something cognitive and internal, a mental space that can be filled or changed in such a way that people can actually be driven to light themselves on fire and sit calmly while letting themselves burn. The realist conception, on the other hand, suggests that cognitively, the homunculus of a man in a burning theater is meanwhile in its own "burning theater," with no choice but to escape and thereby compel the man to escape. We end up accepting the view that, deep down, humans lack even the *cognitive* freedom to choose not to escape.[18]

Now, granted, if someone asks whether people will run out of a burning building, it seems ridiculous to suggest that we have no clue what they might do. But at the same time, just as Bouazizi's self-immolation led to dramatic political consequences, systemic

[17] I would say that genuine, non-selfish cooperation exists, but this doesn't suggest we should never be critical of claims that one is acting selflessly or cooperating. In this sense, social science may usefully reveal situations in which cooperation is a consequence of selfishness (e.g., Axelrod 1981), but this is quite different from the "no true selflessness" dogma.

[18] A related point can be made regarding Niou and Ordeshook (1994).

theorists cannot simply discount what may, empirically speaking, amount to rare anti-system events as inevitably unimportant without importing even more determinism in their theory. As I noted above, what self-immolation represents is that there is no *logically necessary* connection between fire and human behavior.

We might simply say this: neorealist structuralism assumes an "everyone has their price" (for upholding the structure) principle for all actors; not only that, but it assumes the system is always willing *and somehow able* to "pay the price" to maintain itself. There are two problems with this. First, it of course presumes determinism; i.e., humans, at some point along the line, lack the free will to *not* do whatever "it" is. Second, and perhaps more interestingly, the belief that systemic imperatives make a certain outcome inevitable can actually *free* the individual to make even more existential choices, especially if following deontological ethics. For example, one's certainty that a major-party candidate will be elected might free one to vote for a minor-party candidate.

The second problem of hard-socializing theories' tacit determinism builds directly on the first but is ultimately more destructive. I argued above that supposed structural imperatives require humans to be wired in such a way that, e.g., the n^{th} state or individual always fulfills the structural imperative, even if other states have failed to do so. How can a microfoundational deterministic model be compatible with both defections-from and the inevitable upholding-of a hard-socialization theory? That is, how do structural realists understand the court of structure as entertaining Mother Teresas, but always keeping Hitlers on the throne?

Wendt (1987) suggests a point that runs somewhat parallel:

> The principal weakness of a structuralist solution to the agent-structure problem is that, because it cannot "explain anything but behavioral conformity to structural demands," it ultimately fails to provide a basis for explaining the properties of deep structures themselves. [347]

The major weakness of hard-socializing theories, then, is failing to provide a clear microfoundational model wherein individual units disobey the system sometimes but never overthrow it; and not only that, but where human microfoundations could only have led to the specific systemic outcome that perpetuates itself forever.

As stated above, hard-socializing theories can hardly get anywhere without presuming something about the way agents act. Orthodox neoclassical market theory, for instance, makes perhaps often-reasonable but certainly not unassailable assumptions in the law of demand or the law of diminishing marginal utility. Just as robots interacting tells us nothing without knowing something about their programming, we can in no way delimit the social behavior of humans without knowing, or at least assuming, certain things about individual behavior.[19]

None of this is to say that neorealist and other structural theories cannot, for example, concisely describe historical patterns, nor that structural theories *couldn't* be (not to say that Waltz's is) internally coherent models of social behavior. The question is,

[19] Even with selection effects, the problem is we can only have *a priori* knowledge about the nature of empirical selection effects when they are analytic, tautological propositions. So when we make causal assumptions and build definitions such as "states that fail to seek security" as equal to "states that fail to survive," and "survive" is defined as "not persisting to future periods," then of course "states that fail to seek security will not persist to future periods." This ultimately becomes useless without some empirical grounding, and of course the usefulness of empirical evidence is questionable in light of the free will argument.

do they actually tell us anything necessarily true, anything fixed and thereby more accurate than any other internally consistent, matching-the-patterns theory?

Soft-socializing theories: Marxism, structurationism, and social constructivism

In contrast to hard-socializing theories, I define soft-socialization to occur when something about the structure/society/whole influences the actual makeup of individual units; agents/individuals and their interests are constituted in ways by society, and are allegedly not reducible to individuals. Some soft-socializing theories are quite up-front about the determinism they assume at the individual level. Marx (quoted in Cohen [2000]), for one notable example, argued,

> the social production of their life, men enter into definite relations that are indispensable and independent of their will … . It is not the consciousness of men that determines their being, but, on the contrary, their social being that determines their consciousness. [vii–viii]

Consequently, obedient Marxian theories inherit this determinism. On the other hand, modern structuration theories, and notably social constructivism, try to leave room for the agency of individuals while not adopting methodological individualism. Wendt (1987) explains of this view that

> the capacities and even existence of human agents are in some way *necessarily* related to a social structural context—that they are *inseparable* from human sociality. … [Structuration] accept[s] the reality and explanatory importance of irreducible and potentially unobservable social structures that generate agents. [355–6; emphases original]

The focus is on whether structure's interactions with agents occur consciously or subconsciously. Insofar as it is conscious, this paper argues that social constructivism

eventually devolves into the same sort of microeconomic model it so frequently critiques; we get "identities for sale!" in social settings, and structures are really just markets. If agents are constituted subconsciously, this paper will argue that there is an implicit behavioral determinism, and further constructivists must answer why agents, in their consciousness, cannot modify this subconscious constitution.

Wendt quotes Bhaskar (who specifically rejects voluntarism, 361), stating: "[S]ocial structures, unlike natural structures, do not exist independently of the activities they govern" (358), then continues:

> While it may make sense to say that a natural structure has an existence apart from the behavior of its elements, social structures are only instantiated by the practices of agents. The deep structure of the state system, for example, exists only in virtue of the recognition of certain rules and the performance of certain practices by states; if states ceased such recognition of performances, the state system as presently constituted would automatically disappear. Social structures, then, are ontologically dependent upon (although they are not reducible to) their elements in a way that natural structures are not. … In other words, social structures have an inherently discursive dimension in the sense that they are inseparable from the reasons and self-understandings that agents bring to their actions. This discursive quality does not mean that social structures are reducible to what agents think they are doing, since agents may not understand the structural antecedents or implications of their actions. But it does mean that the existence and operation of social structures are dependent upon human self-understandings; it also means that social structures acquire their causal efficacy only through the medium of practical consciousness and action. [359–60]

Wendt (1992) asks, "What in anarchy is given and immutable, and what is amenable to change?" (391). Interestingly, however, while he challenges the ability of structure to dictate/determine the terms with which states choose to face it (viz., self-help):

> [S]elf-help and power politics do not follow either logically or causally from anarchy …. [394]

… he nevertheless permits "process" to serve a similar role:

> There is no "logic" of anarchy apart from the practices that create and instantiate one structure of identities and interests rather than another; structure has no existence or causal powers apart from process. [395]

And he later goes on to say:

> Institutions are fundamentally cognitive entities that do not exist apart from actors' ideas about how the world works … [They] come to confront individuals as more or less *coercive social facts*.

Consider Ruggie's (1998) perspective on how ideas causally connect to human behavior:

> Some ideational factors simply do not function causally in the same way as brute facts or the agentive role that neo-utilitarianism attributes to interests. … Suffice it to say that these [example] factors fall into the category of *reasons for actions*, which are not the same as *causes of actions*. [869]

Ruggie's distinguishing between reasons for actions and causes of actions is illustrated by European integration, where he separates the "*aspiration* for a united Europe" as

> not *caus[ing]* European integration as such, but [being] the *reason* the causal factors … have had their specific effect—in Weber's words, produc[ing] an outcome that is historically *so* and not *otherwise*. Absent those 'reasons,' however, and the same 'causes' would not have the same causal capacity.

Thus, examining all holistic or quasi-holistic theories (Marxism, neorealism, structuralism, social constructivism) very quickly gets us to the question of individual behavior, and the issue of what, if anything, predetermines it.[20]

Taming material

In the preceding section I argued that we cannot rely on structures or societies as a whole to explain human behavior for us without presuming a great deal about how the individual human operates. Explaining social behavior by focusing on societies, or on how social structures dictate or constrain individual behavior, cannot be theoretically sound without denying individual humans' free will. Moreover, I noted that the varying ways in which individuals are thought to operate imply very different things about the importance of studying structures and societies. I then pushed that argument farther to show that at the micro level, even purported "ideational" social-scientific paradigms must be leveraging some material (i.e., biological) behavioral determinant(s). This leaves theories such as social constructivism and neorealist structuralism not only theoretically shaky, but also much less empirically useful.

For a fully aware, logically rigorous social-scientific framework, we must therefore un-bracket the individual and ask whether there are any behavioral determinants at the individual level, and if so, what these determinants might look like and how they will consequently affect the shape of social science. Holistic theorizing such as social

[20] Cohen (2000) gives this concise statement of the microfoundational, analytical perspective: "[Those self-described 'analyticals'] reject the point of view in which social formations and classes are depicted as entities obeying laws of behaviour that are not a function of the behaviours of their constituent individuals" (xxiii).

constructivism may be partially redeemed, but only with a more self-aware position on the determinants of individual behavior.

Thickly rational conceptions of human behavior that suppose some consistent "human nature," on the grounds that such conceptions are unfalsifiable in one sense, easily falsified in another, and only useful in a limited, practical context. Thinly rational conceptions of human behavior are more common as modeling tools based on their purported flexibility.

First, I would like to clarify what I mean by "thick" and "thin" rationality. Consider the definitions provided by McCarty and Meirowitz (2007):

> 1. Confronted with any two options, denoted *x* and *y*, a person can determine whether he does not prefer option *x* to option *y*, does not prefer *y* to *x*, or does not prefer either. When preferences satisfy this property, they are *complete*.
>
> 2. Confronted with three options *x*, *y*, and *z*, if a person does not prefer *y* to *x* and does not prefer *z* to *y*, then she must not prefer *z* to *x*. Preferences satisfying this property are *transitive*.
>
> [O]ur working definition of rational behavior is behavior consistent with complete and transitive preferences. Sometimes we call such behavior *thinly* rational, as properties 1 and 2 contain little or no substantive content about human desires. Thin rationality contrasts with *thick* rationality whereby analysts specify concrete goals such as wealth, status, or fame. The thin characterization of rationality is consistent with a very large number of these substantive goals. In principal, thinly rational agents could be motivated by any number of factors including ideology, normative values, or even religion. As long as these belief systems produce complete and transitive orderings over personal and social outcomes, we can use the classical theory of choice to model behavior. [6–7]

Thick rationality, as defined here, is closely associated with the public choice school of economics, which often assumes, e.g., that politicians are cold-hearted maximizers of "pork" for the home district, or at least of re-election chances. In international relations, this is akin to standard neorealist assumptions that states, and consequently their leaders, are motivated to seek security at nearly all costs.

However, consider the definitions of rationality in Lichbach (1996):

> Rational action in turn may be either thin or thick. Thin rationality means the minimalist definition of rationality offered above—desires and beliefs are internally consistent and consistent with one another. Consistency implies that an actor chooses actions that best satisfy a given set of objectives. The thin version of rationality has been termed instrumental rationality.[21] It has also been called a present-aim theory because it implies the efficient pursuit of whatever aims one has at the moment[.] … Because thin rationality can result in unintended and undesirable consequences, actors may supplement thin rationality with thick rationality or reason. This implies that people can do more than passively respond to choices in a consistent way. They are free to choose, create, and adjust goals; they can also change the situation or constraints in which they find themselves. In short, thick rationality takes the notion that people have intentionality and agency more seriously than does thin rationality. This version of rationality has been termed expressive rationality. It has also been called a self-interest theory of rationality because it implies that a person's goals and beliefs efficiently promote the person's interests. [28]

(Lichbach notes that in terms of rational desires, "[o]ne dichotomy is between self-interest and self-transcendence," and that "material- or outcome-oriented action can focus on pecuniary [monetary] payoffs or nonpecuniary [general political, policy, or moral] payoffs.")

[21] Lichbach here cites Heap, Hollis, Lyons, Sugden, and Weale 1992.

While these authors' definitions are not quite compatible, they are essentially covering the same issues. To McCarty and Meirowitz, thick rationality is associated with "concrete goals such as wealth, status, or fame," i.e., those often seen as objectively selfish, whereas thin rationality is about instrumental or procedural rationality, and lets self-interest include seemingly selfless acts such as promotion of religion. Lichbach refers to thick rationality as *any* of those situations where we can say an actor is consistently pursuing a unified goal, whereas thin rationality is the simple case of consistent preferences in a given setting.

For this paper's purposes, I define thin rationality to mean procedural or instrumental rationality consistent with the axioms of game-theoretic play, i.e., expected utility maximization theory, or a closely related decisional model (e.g., bounded rationality). If a research knows that an actor is thinly rational, he should be able to correctly describe, for any given game, what that actors actions will be give any possible preference. Thick rationality, on the other hand, means both procedural/instrumental rationality *and* the presumption of some objectively "rational" preferences, which in typical usage seems to be *selfish* preferences, i.e., those considered most fundamental in our status as materially/biologically based creatures. Often the thick rationality presumed is in line with traditional assumptions about human nature. Therefore, as McCarty and Meirowitz say, wealth, status, or fame—or power—are likely candidates. In a thickly rational model, given any actor and a particular game, we claim to know the specific action that actor will take, since we know not only that she will be procedurally rational, but also what her preferences are. Lichbach's reference to expressive rationality is more

consistent with thin rationality; but there are some problems reconciling this idea of expressive rationality with traditional thin rationality.

To frame the problem with these models, consider the diner model of causality (far) above. On a related note, Sen (1977) states,

> It is possible to define a person's interests in such a way that no matter what he does he can be seen to be furthering his own interests in every isolated act of choice. While formalized relatively recently in the context of the theory of revealed preference, this approach is of respectable antiquity, and Joseph Butler was already arguing against it in the Rolls Chapel two and a half centuries ago. The reduction of man to a self-seeking animal depends in this approach on careful definition. If you are observed to choose x rejecting y, you are declared to have "revealed" a preference for x over y. Your personal utility is then defined as simply a numerical representation of this "preference," assigning a higher utility to a "preferred" alternative. With this set of definitions you can hardly escape maximizing your own utility, except through inconsistency. Of course, if you choose x and reject y on one occasion and then promptly proceed to do the exact opposite, you can prevent the revealed preference theorist from assigning a preference ordering to you, thereby restraining him from stamping a utility function on you which you must be seen to be maximizing. He will then have to conclude that either you are inconsistent or your preferences are changing. You can frustrate the revealed-preference theorist through more sophisticated inconsistencies as well. But if you are consistent, then no matter whether you are a single-minded egoist or a raving altruist or a class conscious militant, you will appear to be maximizing your own utility in this enchanted world of definitions. [322]

Hence the "mechanistic" processing of utility as characterized by Hollis. The fundamental problem—and strength, many argue—of thin rationality models is the exogenization of interest-formation. But by slicing social interactions into discrete games, each with exogenously defined (and unalterable) rules, players, and the like, this is an

artificial way of creating social structure and determinacy, and falsely suggests the existence of certain "natural" games. Any specific social games that are played are supported by other meta-games that define a game's rules, players, and the like. The meta-games are supported by further meta-games, and the only natural, unsupported game is too vaguely defined to tell us anything about social behavior; see Cartwright (2009) on this point.

Further, even temporarily ignoring problems of utility theory, the grand "Game" for each individual includes the task of inferring or defining a meaning for life, forming beliefs about the supernatural and the afterlife, and so forth. I argue that *contra* Bueno de Mesquita, the lens of "brazen self-interest" is not particularly useful when self-interest can be influenced by all manner of "off-the-path beliefs" in destiny, deities, and the like. I also specifically criticize game theory's common knowledge, complete information, and common priors assumptions as incompatible with twentieth century epistemological developments—such as Quine–Duhem ontological relativity, the failure of logical positivism, and problems in models of knowledge.

One source of identities, interests, and the "rules of the game" is society; but this paper has argued against that above. Only two sources remain: one is that these originate from determinants "below" the human level; viz., biological forces. The other, mutually exclusive source is that, to some degree, these originate via volitional actions.

Carr argued, "The realist analyses a predetermined course of development which he is powerless to change." Mearsheimer (2005) notes, "The fact is that Carr was a

determinist at heart who did not think that individuals could purposely re-order the international system in fundamental ways." (141)

Finally, even if correct in the sense of procedural rationality, utility theories allowing both tangible and intangible preferences that are only known *a posteriori* have little value except as historical descriptions or as illustrations of hypothetical social scenarios. As Bandura (1986) argued, "A theory that denies that thoughts can regulate actions does not lend itself readily to the explanation of complex human behavior."

While utility-theoretic models may be concise ways of representing preferences or prescribing behavioral rationality, they cannot accommodate a number of relevant types of decisions, and even the thinnest version make excessively bold assumptions in order to derive equilibria in games. Therefore humans are free to behave and find meaning in life freely, and this view is existentialism.

The free will debate vs. IR dualities

Another consequence the debate over human free will should have on international relations theory is encouraging more careful attention to what we mean by such terms as "agency," or clarifying what fundamental differences may lie beneath the manifold "logics" of action. Cleaning up our terminology also forces us to confront approaches that try to compromise human volition and imply we have a free-but-constrained will.

I am not the first to make the connection between the volition debate and social-scientific terminology; Wallerstein (1997) explains,

> Macro and micro constitute an antinomy that has long been widely used throughout the social sciences, and indeed in the natural sciences as well. In the last twenty years, the antinomy global/local has also come into wide use in the social sciences. A third pair of terms, structure/agency, has also come to be widely adopted, and is central to the recent literature of cultural studies. The three antinomies are not exactly the same, but in the minds of many scholars they overlap very heavily, and as shorthand phrases they are often used interchangeably.
>
> Macro/micro is a pair which has the tone merely of preference. Some persons prefer to study macrophenomena, others microphenomena. But global/local, and even more structure/agency, are pairs that have passions attached to them. Many persons feel that only the global or only the local make sense as frameworks of analysis. The tensions surrounding structure/agency are if anything stronger. The terms are often used as moral clarion calls; they are felt by many to indicate the sole legitimate rationale for scholarly work.
>
> Why should there be such intensity in this debate? It is not difficult to discern. We are collectively confronted with a dilemma that has been discussed by thinkers for several thousand years. Beneath these antinomies lies the debate of determinism versus free will, which has found countless avatars within theology, within philosophy, and within science. It is therefore not a minor issue, nor is it one about which, over the thousands of years, a real consensus has been reached. [1241–2]

Unfortunately, Wallerstein's treatment of the volition debate is at times puzzling; for example, at one point he claims confusingly, "This is an endless, pointless, sequential chain. Starting with free will, we end up with determinism, and starting with determinism, we end up with free will" (1255). (In fact, Wallerstein's perspective frequently seems closest to hard indeterminacy—i.e., and as noted above, the idea that human behavior is neither predetermined *nor* freely chosen—though this makes puzzling his calls for "utopistics" elsewhere.) For example, he writes later that "The picture of the

universe that derives from this model is an intrinsically non-deterministic one"; but he adds that "it does not follow that the universe can therefore move in any direction whatsoever" and justifies this indeterminacy by saying that "the aleatory combinations are too many, the number of small decisions too many, for us to predict where the universe will move." In his justification (i.e., his emphasis on "aleatory combinations") his perspective on debate would seem to lie closer to the compatibilist view of Dennett (2003). Elsewhere, upon concluding a passage that fits closely with the arguments elsewhere in this paper, he declares—in contradiction to this paper,

> [W]e must stop fighting about non-issues, and foremost of these non-issues is determinism versus free will …. For very long and very short time spans, and from very deep and very shallow perspectives, things seem to be determined, but for the vast intermediate zone things seem to be a matter of free will. We can always shift our viewing angle to obtain the evidence of determinism or free will that we want. [1255]

Meanwhile, Hollis (1983) also notes in his critique of "Rational Economic Man" that "[i]ssues of individualism *vs.* holism, of psychologism vs. sociologism, and of choice vs. determinism loom" (253). Doty (1997) echoes this, arguing, "The controversy over the role of agency versus that of structure in social life has touched nearly all areas of the social sciences, reflecting longstanding concerns with and attempts to come to grips with several powerful dualisms—holism vs individualism, objectivism vs subjectivism, determinism vs voluntarism" (365). And Carr argued,

> The antithesis of utopia and reality can in some aspects be identified with the antithesis of Free Will and Determinism [*sic*]. The utopian is necessarily voluntarist: he believes in the possibility of more or less radically rejecting reality, and substituting his utopia for it by an act of will. The realist analyses a predetermined course of development which he is powerless to change.

Just as others have recognized the important underlying connection the volition debate has to more commonly discussed social-scientific concepts and dichotomies, this paper takes the perspective that free will and determinism may help "mop up" some of the definitional variance of many commonly used IR dualities.

Thus, assuming, as this paper has from the outset, that determinism and free will are incompatible viewpoints, then the determinist vs. libertarian/voluntarist tension may have several analogues in IR theory:

- *Macro (structure/global/whole/society/culture) vs. micro (agent/local/part/individual/person)* — As has been argued above, it becomes very difficult to reconcile straightforward "macro" models with a voluntaristic outlook, which emphasizes the micro. Deterministic perspectives may focus on either macro- or micro-level explanation, although (as will be discussed below), the question remains what work is being done according to either macro- or micro-level causal mechanisms in a determinist's IR model.
- *Objective vs. subjective* — Because determinism implies that the causal chain of human actions is only experienced, not mediated, by human consciousness, the distinction between objective and subjective becomes less relevant. Subjectivity and objectivity are immaterial are effectively indistinguishable at the level of causality in the determinist's world, where human behavior is a consequence of predetermined neurological events. A voluntaristic perspective clarifies the distinction between objective and subjective, but also

renders the distinction irrelevant. Humans act with respect to either objective or subjective truths, but the relationship becomes entirely open and indeterminate.

- *Positive/realist vs. Normative/utopian* — The classic dichotomy between positivistic explanation and normative exhortation becomes sharper in the shadow of the free will debate. Without free will, social scientists could have no normative role (in the sense of possessing the ability to change future outcomes for themselves or for society); only documentation of existing causal relationships is possible.[22] On the other hand, in a world in which humans have free will, social scientists cannot point to any causal relationship in human affairs as *necessarily* true because of the intervening indeterminacy of free will, so the role of normative appraisal becomes more important. To Carr, this boils down to the debate between realists and utopians, with the latter group naively ignoring the unyielding causal forces inherent in human nature and social affairs.

- *Logics of action* — The manifold "logics" of action commonly appealed to in IR literature—such as consequences, appropriateness, practice, and habit—do not align neatly with determinism or voluntarism; but that latter dichotomy may usefully apply to and clarify the role of the logics. Of course, each of the logics is reconcilable with determinism in that social scientists have used the logics to explain causality: e.g., human agents deterministically act in a certain way because of a certain logic (or illogic, as may be the case with practice and habit). With the former two logics, a voluntarist perspective sees a *choice* of grounds an

[22] Granted, the obvious point is that to a consistent determinist, *whatever* social scientists do, consistent with their philosophical assumptions or not, they could do nothing else.

individual uses to make decisions, whether by consequences or appropriateness, while with the latter two logics, the voluntarist sees one's ability to temporarily relinquish free will by relying on practical or habitual behaviors.

- *Material vs. idea(tiona)l* — Finally, the difference between material and ideational explanation is again clarified in context of the free will debate. While a determinist may casually think of ideas as causal, as the next section argues, he must admit that if determinism is correct, then underlying even "thoughtful" action is the raw material force of biology, physics, etc. On the other hand, for the voluntarist, while material may constrain choices or provide incentives in favor of certain choices, the nature of free choice is that it is always subject to a cognitive, ideational process whereby ideas—objective or subjective—may come into play.

Thus, the usefulness and coherence of these dualities changes in light of a clear-headed, consistent perspective on the free will debate.

Chapter 3: Understanding Free Will with Applications for International Relations

Given the assumptions and definitions this paper has given, there are only two coherent perspectives on human possession of free will. As emphasized in the foregoing section, this implies that IR theorists are left with only two choices as to how the volition debate impacts and interacts with international relations theory: all theories, whether "big picture" or more specific, ultimately owe allegiance, implicitly if not explicitly, to one of those two perspectives. IR theorists must accept determinacy and build theories and models with a purely deterministic approach, or, if a theorist begins with the proposition that humans have free will, he must build theories and models concordant with a libertarian approach. (Because of the latter's alternative political meanings, this will be termed an *existential* approach.)

The following subsection will first briefly reiterate the significance of a deterministic view of international relations, then turn to outline the possibilities for an "existential turn" in IR theory. The remainder of the paper continues on the premise of such an existential turn.

Determinism in IR theory

There is a contradiction inherent in the deterministic approach to IR theory, however. If we are confident determinists, then to build causal/explanative (non-descriptive) social science, we must start by accepting the deterministic outcomes in the domain of natural science that constrain human behavior: both at the biological level (the genetic, epigenetic, phylogenetic, and environmental determinants of human behavior) as well as the physical/chemical level.[23] But if we accept these forms of explanation, then traditional social-scientific becomes purely epiphenomenal at best and fantastically incorrect at worst, hence the puzzle. A determinist who constructs a social science theory intended to determine human outcomes, even within a range or probability distribution, must ask herself what work the theory's purported causal mechanisms are doing in addition to underlying biological and chemical mechanisms.

Indeed, in examining such models closely, human cognition, understanding, and experience of social structures, culture, norms, incentives, and the like, would seem to be a mere subjective veneer over top of the actual causal story doing the work on a physical level. Even if the causal mechanisms at work are neurological and hence experienced subjectively, the language frequently used to describe these mechanisms—e.g., "The decision-maker chooses *x* because of incentive *y*"—would remain false (or, at least,

[23] And perhaps at other "levels" as well, including the supernatural level depending on one's philosophical outlook; for example, religious adherents may view humans as having the *potential* for free will (in a natural sense), but for many humans' actions being constrained by supernatural factors.

disingenuous) in the use of terms such as "decision," "choice," and so forth.[24] Even if one argues that the use of such language is acceptable from a word-use standpoint, it becomes greatly confusing in making apparent normative or policy recommendations.[25]

All traditional social science thereby becomes interpretive, an "inside" look at what are, essentially, illusions inserted into our consciousness by subcognitive and other external structures.[26] Aside from practicing this sort of social scientists, international relations theorists could give up on explanation and try to craft predictive social science based only on historical data—presuming, likely with some Bayesian basis—that human behavior will tend to look like it has in the past, and hence, humans will probabilistically behave in certain ways at any point in time. While useful, this both represents a resignation that we cannot understand human behavior, fundamentally, and also leaves social science as something akin to statistically rooted historical inquiry. This means we cannot actually understand the form human action will take in the future.

Thus, the alternative to free will is not a sensible social science; social science withers away as epiphenomenal, our conscious experience of subcognitive physical states is an illusion, and our relationship with what actually determines our thoughts and actions remains enigmatic. To continue the analogy from above, if what goes on in our minds is but another domain in which, were we omniscient, we would see the equivalent of billiard-ball causality, then social-scientific causality is really no different from natural-

[24] The compatibilist view may offer a different perspective here.
[25] Two things. First, explain how this can quickly lead us into the "black hole" of thinking about how, well, why are we doing social science if we can't change anything, but then, hey, we can't even change our doing of social science. Second, how it's different from the equivalence theorem for natural science b/c it presumes we do have free will, so, e.g., understanding heart disease makes a difference
[26] And if nature is fundamentally indeterminate, even the hope of a "complete psychology" disappears.

scientific causality. If this is true, the way we understand and interact with the world is predetermined; our level of understanding of the world, even if true, is predetermined, and we are certainly powerless to alter its workings; and social science becomes completely subservient to natural science in actually explaining the origin, progress, and ultimate outcome of any social situation.

A final note: what are the consequences of building deterministic models and theories if determinism *isn't* true? First, consider the legal import: the function of law is generally premised on the notion that, except in specific cases, lawbreakers could have acted differently, viz., in accordance with the law. The U.S. Supreme Court declared in 1978:

> The Scott rationale rests … on a deterministic view of human conduct that is inconsistent with the underlying precepts of our criminal justice system. A "universal and persistent" foundation stone in our system of law, and particularly in our approach to punishment, sentencing, and incarceration, is the "belief in freedom of the human will and a consequent ability and duty of the normal individual to choose between good and evil."[27]

Apart from this basis, issuing any sort of genuine moral opprobrium becomes unfair and absurd, like scolding a volcano that it shouldn't have erupted. More troubling, there is a performative aspect: belief in determinism actually *abets* unethical behavior (Vohs and Schooler 2008; cf. Viney et al. 1982).

[27] United States v. Grayson, 438 U.S. 41 (1978). See also Burns and Bechara (2007), Cotton (2005), O'Hanlon (2008), Sasso (2009), and Atiq (2012).

Existentialism in IR theory

The opposing view, in the extreme, is no less unorthodox. If humans possess free will, then at some fundamental level, an individual has total autonomy in his or her free actions from not only society, but also from any obvious material or supernatural constraints, and even autonomy from reason (especially because humans lack any undeniable epistemological foundation), we lose any solid basis for theoretically grounding the causes human action.[28] This essentially (pardon the pun) existentialist perspective is summarized by Sartre's (1998/1943) arguments:

> [M]an first of all exists, encounters himself, surges up in the world—and defines himself afterwards. If man as the existentialist sees him is not definable, it is because to begin with he is nothing. He will not be anything until later, and then he will be what he makes of himself. Thus, there is no human nature .... Man simply is. Not that he is simply what he conceives himself to be, but he is what he wills, and as he conceives himself after already existing—as he wills to be after that leap towards existence. Man is nothing else but that which he makes of himself. … For if

[28] We might conceivably, but ultimately unhelpfully, reconcile existentialism with rationality in two ways. First, if our behavior is predetermined, then the perception of existential choice is an illusion. Nevertheless, instances of existential behavior remind us that the manner in which our behavior might be predetermined must be compatible with absurdity and off-the-path choices "for the sake of off-the-path choices." Therefore whatever determining mechanisms exist at the cognitive and subcognitive/neurological levels are *not* the same as traditional human nature / thick rationality assumptions. Those who want to rely on those assumptions, then, must do one of two things. They may answer the question of why we should believe the *real* mechanisms that determine our behavior will systematically lead to something looking like thick rationality, despite evidence of defections. (One possibility could be that the ordinary individual simply lacks the mental computing power to reach a state of existential crisis or have true freedom of action.) Or they may simply redefine "rationality" to mean "whatever our biology compels us to do," which, given the range of human behavior, "rationality" becomes tautological and useless. Nevertheless, if existentialist behavior is generated deterministically, then even without the possibility of free will we still are left with a random-walk model of individual behavior when looking at the social level; only by digging into biology can we understand why humans behave rationally, irrationally, selfishly, selflessly, absurdly, etc. Existential behavior is nearly as enigmatic as free will. While it leaves scientists one "escape route" for finding determinacy, viz., subcognitive determinacy, it still wreaks havoc with traditional social science. Further, if the subcognitive escape route is either untrue or unverifiable (or if the universe is fundamentally indeterminate), the specter of free will remains, and it becomes difficult to identify what is irrefutably part of human nature.

> indeed existence precedes essence, one will never be able to explain one's action by reference to a given and specific human nature; in other words, there is no determinism – man is free, man is freedom. … [T]o begin with [man] is nothing. He will not be anything until later, and then he will be what he makes of himself. Thus, there is no human nature …. Man simply is. Man is nothing else but that which he makes of himself.[29] [488]

Along those lines, Weber (1897/1956) famously wrote, "[W]e are cultural beings, endowed with the capacity and the will to take a deliberate attitude toward the world and to lend it significance" (180); Penn (2011) writes, "People take an active role in interpreting their own experiences, and in assigning meaning to those experiences" (2); and McIntosh (1995) points out that even imagination and false beliefs can lead to "propositional acts." One encyclopedia's definition of existentialism explains that

> human beings are not solely or even primarily knowers; they also care, desire, manipulate, and, above all, choose and act. … [M]an is not a detached observer of the world, but 'in the world.' He 'exists' in a special sense in which entities like stones and trees do not; he is open to the world and to objects in it. … The claim that man exists in this unique sense also means that he is open to a future which he determines by his choices and actions; he is free. … *[N]either as a species nor as individuals do human beings have such an essence that governs their conduct. Man makes himself what he is by his choices, choices of ways of life … or of particular actions …. Even when he seems simply to be acting out a 'given' role or following 'given' values—given, for example, by God or by society—he is in fact choosing to do so, for there are no given values that can determine, in and of themselves, rationally or causally, man's choices*. It does not follow that the

[29] Somewhat similar is Gould (1988): "We are here because one odd group of fishes had a peculiar fin anatomy that could transform into legs for terrestrial creatures; because comets struck the earth and wiped out dinosaurs, thereby giving mammals a chance not otherwise available (so thank your lucky stars in a literal sense); because the earth never froze entirely during an ice age; because a small and tenuous species, arising in Africa a quarter of a million years ago, has managed, so far, to survive by hook and by crook. We may yearn for a "higher" answer—but none exists. This explanation, though superficially troubling, if not terrifying, is ultimately liberating and exhilarating. We cannot read the meaning of life passively in the facts of nature. We must construct these answers for ourselves."

> choices available are unlimited. His 'being in the world' implies that man is 'thrown' (Heidegger) … into a specific situation, and not all the choices that that seems to leave open are in fact possible; but which ones are possible and which are not cannot be known in advance. … [Existentialists] have argued that the openness of the future and the specificity of individuals and of their situations elude rationalist philosophical systems.[30]

According to that definition, note that man is seen to determine, or cause, his own choices, matching exactly with Inwagen's discussion of immanent or agent causation as defining free will (discussed above).

Existentialism provides a way of paradoxically understanding—by not understanding, in a sense—individual behavior as essentially indeterminate. Furthermore, for Kierkegaard and others, existentialism is the pathway to the absurd. This reaffirms the comments on ontological relativity above; not only are we epistemologically ungrounded, but the premises of existentialism are that even where there might be knowledge (or existing social behaviors, roles, identities, etc.), it stands separate from action; no knowledge or history compels, necessitates, or determines human action.

Existentialism thereby represents a total rejection, both positive and normative, of the idea of a pre-set human nature. Accepting the "existence precedes essence" dictum threatens social science as commonly practiced, because in any social institution or situation, actors may select any form of behavior, any strange (e.g., "absurd") or off-the-path preferences, etc. There is no necessarily stable "mean" of human behavior and therefore no stable mean of institutions; and if ever we should see one we can only say it was of our own volition, and could disappear at any time.

---

[30] "Existentialism" in *Encyclopaedia Britannica*, 15th ed. Emphasis added.

If our behavior is free, the game theorist might argue that he could, theoretically, reconcile such free, "existential" behavior with game theory by postulating that an individual endows an existential choice with utility simply by virtue of the fact that she believes she is making a free, existential choice. There are three problems with this. First, it requires understanding the nuances of an individual's existential behavior at near-*verstehen* levels, or at the very least, a closely situated understanding of the peculiarities and idiosyncrasies of an individual, necessitating their honesty. Performativity—i.e., if the individual wanted to elude rational description—could also creep in. Second, the existentialist's behavior would, at best, only be consistent with game theory in between acts of will, meaning a researcher would need to be something like best friends with each research subject in order to understand their behavior well enough to model it; but then, what's the use of game theory if you are so close to someone? Third, any form of game theory, even if it can accommodate existentialized *preferences*, necessarily presumes instrumental rationality, but existential acts may violate such rationality. Therefore modeling existential behavior as thin rationality likewise is tautological and no more useful (and far less readable) than a simple biography.

Having argued that both structurally and materially determining IR models cannot easily grapple with the possibility of free will, this paper now turns in a new direction: examining the implications of the existence of free will for political science. Specifically, how can political scientists usefully conceptualize free will as a component of the humans they study, and how does this change the nature of political science?

Voluntarist theories of international relations

Previous sections laid out the argument for why the existence or nonexistence of free will is no trivial matter for social science in general, and international relations theory in particular. While this paper attempted to show the consequences the free will debate has for specific theories and paradigms in social science, it now turns to a broader question: what would the presence of free will mean for the practice of the international relations field of political science as a whole? That is, supposing this paper has successfully defended my claim that this problem necessitates a fundamental shift in social theorization, may we move forward, and if so, how?

On this front, Adcock (2007) writes,

> A clear-headed embrace of any of the principal philosophical species of indeterminism would call for shifts in explanatory practices so significant as to effectively terminate the tradition of macro-historical inquiry as pursued by American sociologists and political scientists during the last half century. [347]

Drawing on Searle (1995), a voluntarist perspective maintains that, in the realm of social facts, there is only agency, and hence society as a whole has the "freedom" to create whatever social facts can logically exist at any time.[31] This does not mean that there are not processes by which society ends up acting as a "whole," only to mean that these processes both descend from voluntary acts and are all continually subject to instantaneous change based on free human choices.

[31] Social facts are things that "exist only because we believe them to exist" (1). While Searle does reference "structure," the context refers not to a structure of causation as in structuralist theory, but rather simply the vast milieu of social facts, especially those taken-for-granted, than enable modern life.

In the absence of uniform, genuine causation in the social realm, i.e., general "covering laws" of social science, or cumulative social-scientific research—in other words, in the voluntarist's world, what should social scientists *do*, practically speaking? This paper offers not a single, perfect model, but rather several overlapping examples of the sort of social study that aligns with a voluntarist perspective.[32]

Oddly enough, a perennial topic in the philosophy-of-social-science debate concerns our similarity to the natural sciences, but if some of the points of this paper are correct, social "science" may be more similar to engineering. Specifically, contingent on our level of free will is the relative importance of studying what *is* versus what *can be*.

First, the voluntarist perspective is reasonably compatible with the model presented in Flyvbjerg (2001, 2004) and "phronetic" social science in general (Flyvbjerg et al. [2012]). This vein of work might be called "problem driven" in the extreme: focusing not even on the less-abstract problems of mid-level theory, but directly on the day-to-day problems of policymakers and other "lay" practitioners, as well as normative and value-rational questions about managing society. Methodologically, it may frequently be considered Feyerabendian. This emphasis is necessary (1) because social behavior is indeterminate and poorly, or at least imperfectly, predictable (even over the short run) in a voluntarist world; (2) because understanding the specifics of a social situation requires understanding, in whatever way it is possible, the intentions of individual actors, even if those intentions are not causal in a determinate sense; (3) because there is a recognition that social behavior is a weaving-together of behaviors

[32] Consider Flyvbjerg (2001, 2004), Flyvbjerg et al. (2012), Kratochwil (1989), Onuf (1989), and especially Price (2008).

grounded in norms, yet, again, actors do not deterministically follow norms and, indeed, may be actively searching for new, better developed, more nuanced or sophisticated norms that resolve contradictions in or inadequacies of old norms; and because (4) insofar as political and social problems interact with technological problems, practitioners frequently need to grasp important elements from both classes in order to assemble optimal solutions and ensure social behavior has a good "relationship" with the material world. Hence, social scientists (to be redubbed shortly) are something like public-sector technocrats equivalent to operations officers and logisticians in the private sector.

Second, the key role of this sort of social scientist might be to "invent" new norms, new practices, new social structures, new identities, and the like for the sake of solving practical problems. In this sense, social scientists might be thought of as social engineers, and acknowledging the presence of some negative connotations with that term this paper will nevertheless use it henceforth. One distinction separating social engineering from the typical application of social science to policy is that the products of social engineers are designed not as once-and-for-all units of knowledge, but rather as makeshift, contingent, temporary social creations that can work no better than the people who follow them and that often fail. Thus, just as a mechanical engineer may recognize the need for emergency backup system to keep a main unit operating, the social engineer must constantly consider the host of things that could go "wrong" with a system and decide what appropriate reactions to each such situation are. Moreover, the engineer must deftly remain aware of how each unit in a system is operating; just as a mechanical

engineer may worry about a particular part failing and must undertake efforts to keep it working, a social engineer must, in a sense, be a psychiatrist or counselor with respect to each individual in a social group, ensuring that the individual play their roles properly in a social system.

While social engineers must be deployed close to problems, this is not to forbid social engineers from designing grander projects. Social engineers can still tackle whole-society problems or even propose utopias; this is close to the call for "utopistics" from Wallerstein, who writes,

> The possible is richer than the real. Who should know this better than social scientists? Why are we so afraid of discussing the possible, of analyzing the possible, or exploring the possible? We must move not utopias, but utopistics, to the center of social science. Utopistics is the analysis of possible utopias, their limitations, and the constraints on achieving them. It is the analytic study of real historical alternatives in the present. It is the reconciliation of the search for truth and the search for goodness.
>
> Utopistics represents a continuing responsibility of social scientists. … Analyzing structures does not limit whatever agency exists. Indeed, it is only when we have mastered the structures, yes, [*sic*] have invented "master narratives" that are plausible, relevant, and provisionally valid, that we can begin to exercise the kind of judgment that is implied by the concept of agency. Otherwise, our so-called agency is blind, and if blind it is manipulated, if not directly then indirectly. We are watching the figures in Plato's cave, and are thinking that we can affect them. [1256]

Wallerstein also quotes Prigogine (1996), who writes, "The possible is richer than the real" (83). In a world of individual free will, individuals undertaking free action may rely on either pre-existing rules/norms or imagination in order to act freely, and perhaps

ideally or even necessarily they possess both and use both simultaneously. We might expect humans to frequently act in such a way that their actions fit into a picture of how the world should be (whether objective, intersubjective, or subjective, and whether ideal in the utopian sense or purely selfish). Even in the voluntarist world, it seems that none of us can be freer than we are able to imagine, nor can our actions be truly free without being individual or social "inventions" in some sense. Put another way, without intentionality, action born out of indeterminacy is merely noise.

The point, in the context of social engineering, is that both individual-level decisional rules and norms and social-level rules and norms (of constitution/convention, etc.) are for social engineers to invent and produce, not for social scientists to discover.

Wallerstein also refers to "analyzing" and "master[ing]" structures. This paper takes the view that such analysis and mastery refers not to causal assertions such as, e.g., how states behave under anarchy, but rather to the normative-laden task of evaluating what outcomes a particular social arrangement might produce: for example, the equivalent of determining what the net production of widgets would be based on a certain arrangement of factory production. The social engineer may still prefer to draw the blueprints for (e.g.) egalitarian societies, or societies that produce ample food, or whatnot. Presumably, social engineering may want to fulfill as many socialized material wants without require excessive exertion of free will, and trying to do so constitutes practical constraints for social engineers.

Thus, free will places social science in a deeper conversation with ethical philosophy and religion than with the natural sciences; voluntarism extends some of the conclusions of constructivism, as in Price (2008).

One reason for social rules is that people (or bands, such as Ten Years After, famously) often seem to cry, "I want to do something [about a social problem], but I don't know what to do," or, "I didn't realize my behavior was causing social ill." In other words, contrary to the simple-selfishness thesis, people sometimes act selfishly simply because they are either unaware of or unsure how to practically deal with a real social problem, or because the complexity and size of modern society make inadvertent deleterious effects more likely. Thus, in line with the critical project, social engineering is about pointing out problems with our social structures and showing people through reintroducing old norms, introducing new ones, or demonstrating on the small scale how to avoid or solve problems. An example from international relations might be Fearon's (1995) suggestion of states using side payments, alternation, or randomization to resolve problems of rule. The practical details of alternation (Who does what when? Why should we follow this procedure? What happens if we disagree in the future?) would be central to a social-engineering approach, and big-picture theoretical work (e.g., in mechanism design) is more a toolbox for the social engineer to play with than a one-size-fits-all way to solve any social problem.

On this point, social theorists' work on deliberative democracy (not the communicative action aspect, but the information transmission aspect) also represents an exemplar. The connection of normative theory and understanding material constraint is

understanding how individual humans may transmit information. Social engineering suggests that we should think about the logical limits of human behavior (and, hence, society): what rulebooks could *not* form the basis of a society? What programs would output the social equivalent of a compile-time or run-time error? For example, a social engineer who devised an agrarian utopia without providing for individuals who can communicate what crops should be grown where and when might clearly fail without normative rules about what to grow, or how to behave in the face of famine.

Along these lines, psychologists and neuroscientists play an important role in light of social engineering. As discussed above, utopias must have layered normative structures for handling unforeseen contingencies, and the golden rule is an ideal fail-safe. But the workability of normative structures, and our free will in accordance with norms, is constrained materially by the ability of the human brain to keep track of what is appropriate at any given time. A utopia with a thousand-page normative rulebook and no easy way to keep track of the rules and the codes for when they apply is likely infeasible, no matter how free our wills are, because of our neurological inability to keep track of and accurately apply the rules. On the other hand, a utopia with a single page of rules may be infeasible as well, because it is unlikely a single page of rules could encompass the diversity of behavioral norms for anything but an extremely simple, and therefore likely suboptimal society. Therefore a supreme job of utopistics and social engineering is to understand the relationship between, and craft, optimal rulebooks and optimal

societies. (Technology may also be of service here.) Further, the "clarity" of a utopian description is important; vague utopianism is like a sloppy blueprint.[33]

Third, the combination of the moral/normative role of social engineering and the need for up-close-and-personal interaction with individuals that a social engineer is working with suggests the roles of personal counselors and psychiatrists, as well as religious and ethical leaders. The social engineer must engage in an honest dialog with both the individuals in society and a community as a whole, both as servant and as leader. In this sense, of course, political and community leaders are frequently the closest we get to social engineers. Social engineers can try to reinforce identities or change them, remind individuals of their long-term goals to strengthen them against short-term aberrations, develop practical solutions to resisting various forms of temptation, and invent unique heuristics to help individuals behave "rationally" in specific situations, with "rationality" having a potentially unique meaning for each individual. Nevertheless, the social engineer should not lose sight of the normative social context in which the individual wants to exist.

Of course, counselors deal with groups of people, too, and conflict management is a role as well. Here we see the ideal type of a wise judge executing Solomonic decisions designed to uphold existing rules, placate parties (sometimes by appealing to their normative convictions rather than their selfish demands), and establish sound precedents.

Thus, the ideal social engineer looks something like a personal counselor or spiritual leader; something like a team leader, manager, or coach; something like a

[33] The implication also seems to be that more complex societies carry an inherent cost in our decreased cognitive ability to understand and optimize them.

project consultant; something like a wise and clever judge; something like a good carpenter or plumber; something like an economist skilled in mechanism design or the applied political scientist.[34]

The ideal social engineer looks something like an anecdote-filled historian, too.[35] Good social engineers are leaders with the desire, brainpower, and tools to invent norms and actions, propose compromises, etc., to maximize everyone's utility and achieve society's goal equilibria.[36]

And actually, almost all social science as practiced today remains somehow useful in a voluntarist world, provided it adjusts its meta-theoretical position accordingly. Grand

---

[34] Mechanism design is a good example for why I am fond of terms such as "engineering" and "invention." Much of auction theory, while seemingly distinct in its mathematical sophistication from practical, lay-oriented phronetic social science, nevertheless has the flavor of taking a problem and solving it using a novel form of social structure. Thus, the Vickrey auction was not a practice that social scientists merely discovered; it was a socially structured economic setting designed for a particular purpose, with particular moving parts, with a particular outcome, and with particular advantages and disadvantages. Further, individuals must choose to act in specific ways for it to work correctly. It even rests upon some norms (although nothing more than the normative background most microeconomic theory presumes, e.g., by failing to model theft or robbery as options). In this sense, our ability to create auction formats that generate efficient outcomes is a good example of social engineering. Of course, auctions do not delve into issues of preference-formation. Social "science" is thus a process of discovery/development of mechanisms, akin to computer programs; the actual "science" underlying the engineering aspect is more like computer science than natural science; we more design programming languages than discover what nature has given us (but, just as with computer science, we must be mindful of underlying, but nondetermining, material limitations and constraints.)

[35] Given free will, knowledge of history is still needed, much out of respect for Santayana's dictum about repeating it. It is not that historical scenarios deprive us of free will; rather, an individual must understand something about his situation in order to exercise free will. For example, the first human to drink fermented grape juice may have thought, "Oh, this a nice feeling, let's have some more!" without recognizing the effects it would have; consequently, this human unwittingly compromised his free exertion of will. This is why social engineers should be cognizant of material constraints and should be armed with backup plans, emergency measures, stopgap mechanisms, and the like. Further, this line of thought suggests that human history could be a path toward *more* freedom as we learn more about the consequences (especially material, but also social) of our actions. We have had enough experience with utopias to know that we can't be naive about how easily we can ignore our biological urges. But just as repeated failures achieving controlled flight (failures we often laugh at as foolish today) did not stop the Wright brothers, specific failed utopias should not wreck the general utopian vision.

[36] In zero-sum situations this runs parallel to the argument that war requires collective intentionality. In a true zero-sum situation, if everyone thinks they would rather fight than compromise, and decides fighting and losing is preferable to not fighting (i.e., the net utility from war is positive irrespective of the expected outcome), then in the absence of a clever compromise, war *is* the best outcome for society.

theory becomes ideal-typical modeling; empirical hypothesis-testing becomes historical inquiry, and so forth. But it all must adopt a critical spin, albeit one that emphasizes not only the unnaturalness/contingency of our current social structures, but also the freedom humans have to create, interpret, act, feel, and (more than anything) to experience and exert free will. It is an existentialist social science, because if individuals are free, social engineering can do no more to influence individuals than any other ideas or social constraints can, and even well-reasoned, utopian schemes can fail no matter how perfect they are in theory. Moreover, social scientists and social engineers are free to define their work and projects as they see fit. (For example, "utopistics" could be latter-day Manhattan Projects, because we cannot predict the consequences of what we're doing.) The purpose is to think of mechanisms, rules and rulebooks, and computer programs for society that enable free individuals to create and sustain any logically coherent society.

Nevertheless, in the face of a voluntarist social ontology, most current social-scientific efforts seem strange. Today, the gold standard of a social scientific theory (especially in political science) is its ability to correctly map onto historical data, whether qualitative or quantitative. Imagine the state of technology if engineers were preoccupied with reverse-engineering yesterday's technology rather than creating new designs for tomorrow! As Wendt (2001) states,

> Positive social scientists are after "explanatory" knowledge, knowledge about why things happen. This is necessarily backward-looking, since we can only explain what has already occurred, although there is the hope that with good explanations we can predict the future. Policymakers, and institutional designers, in contrast, need "making" or "practical" knowledge, knowledge about

> what to do. This is necessarily forward-looking, since it is about how we should act in the future. As Henry Jackman puts it, "we live forwards but understand backwards." The former cannot be reduced to the latter. Knowing why we acted in the past can teach us valuable lessons, but unless the social universe is deterministic, the past is only contingently related to the future. Whether actors preserve an existing institution like state sovereignty or design a new one like the EU is up to them. The voluntarism inherent in this question is something that positive social science is not well-equipped to handle.

On the morning of the social equivalent of a tic-tac-toe game, a social engineer who knows the players, and knows they want to win, can use that knowledge, and the knowledge of the game's rules, to think about the outcome. But voluntarism suggests that the social equivalent of a tic-tac-toe game is not natural; it isn't something we can predict will occur 1,000 years hence; there are an infinitude of slight and not-so-slight variations of tic-tac-toe that might occur. Further, the players in those tic-tac-toe situations may be bitter rivals, but they also may be parent and child. The normative background and associated rationality that will give the tic-tac-toe game (the institution) its "equilibrium" character cannot be known ahead of time. The usefulness of something like game theory, then, is highest in between specific exertions of will, and virtually nil and noncumulative over the long haul.

In the end, big-picture social engineering becomes something like science fiction. Just as a science fiction writer may introduce a new technology (without actually knowing how to build it) that gives a real scientist or engineer the inspiration to develop the practical blueprint for the technology, big-picture social engineering is like saying, "Here's what we want, and why we (should) want it." Phronetic engineers can come

along, inspired by the big picture, and figure out how to implement it in the real world, borrowing from and contributing to a practical toolbox of social and policy practices. But day in and day out, the maintenance of that technology requires handy mechanics, parts that work properly, and the like. Nevertheless, social science needs more visionaries, more problem-solvers, as in engineering.

In responding to Hollis and Smith's (1990) meta-theory of international relations,) "wish[es] they had said more about how we might study processes of role-structured reasoned judgement" (385). If such judgment is free, then a question arises as to if, and how, we could ever study or seek to improve our understanding of free will. Also, paralleling the question Tullock (1981) asked in response to game-theoretic revelations that there was no "core" in preference aggregation models, we might ask "why so much stability?" If we look at the real world, we seem to see a great deal of structure and constraint, and only occasional acts of agency and freedom. How does free will actually "fit onto" the material world? In reference to my billiards metaphor for free causality, what is the equivalent of the cue, i.e., the interface between free action and mental causation?

My discussions above of homunculi and metapreference structures remind me of those feeble attempts. Even while acknowledging Sartre's view of a human as self-defining, this paper does not deny that humans are pulled—in all directions—by their often contrasting biological urges, social calls from all perspectives, and so forth.[37]

[37] Biological urges can contrast as when one wants to simultaneously maximize life expectancy, maximize reproductive possibilities, and maximize security. Guns, butter, or concubines? Material imperatives are by no means straightforward all the time.

What's more, I recognize some form of causation that occurs between socialized material imperatives/constraints and our behavior, even if, like the restaurant patron's decision described above, the causation is not logically necessary.

Thinking about the mainstream psychologist or social scientist's plight in trying to piece together all the causal forces and come up with some way to delimit human behavior, there seems to be a similarity between such attempts and the epicycle-riddled astronomy of geocentric days. With enough epicycles, we can—tautologically—"explain" any particular behavior. (And the epicycles do have some validity to them; again, I am not denying the influence on human behavior of society, biology, etc.) When we speak of accounting for variance with a mix of realist, institutionalist, and constructivist theories of IR, we are really building epicycles-on-epicycles, and arguing which is the actual deferent.

But at some point we need a frame-shift. Replacing our current social ontologies with voluntarism is a big enough task, but understanding how the exertion of will relates to existing pressures is perhaps bigger still. Perhaps the billiards table is again useful (although this doesn't fit with my reference to cues). Could we arrange the billiard balls in a particular pattern if we all agreed? Yes. Do we each have free will in striking the billiards table? Yes. Are there still impediments to creating any particular arrangement of balls with any particular shot? Yes. Does the present location of balls influence our shots? Yes. When we strike a ball freely, are there, sometimes, inadvertent effects? Yes. When we strike a ball, does the ball keep rolling, sometimes farther than we meant it to?

Yes. Can we wave our cues in the air in any pattern we like, without constraint by the billiards table walls? Yes.

A voluntarist social ontology means that we recognize voluntarism in the social "yard," but it does not necessitate the rejection of determinism in the biological/natural yard. Moreover, our bodies are built to behave deterministically in many circumstances (e.g., in one's reflexive reaction to touching a hot stove), and, as assumed above, we can freely "turn off" our free will in certain situations (e.g., attempting to suppress any higher reasoning about our actions). So we are free to move the boundary between the free-will yard and the determinism yard back and forth, even claiming it's all deterministic territory if we would like. But when it comes to humans, the part of our behavior that has the potential to be free can never follow deterministic laws in perpetuity, because it each individual decides where the boundary is, and because every time we give free will a chance, we change the landscaping of the free will yard enough that when the deterministic boundary creeps in again, the predetermined behavior will not necessarily be the same at all as last time. Wild grasses left alone will result in something quite different than Zen gardens left alone.

Norms, rules, guilt, and critical theory not only can empower individuals to make better choices by identifying what to do in complex social situations—much like shorthand solutions for cooperative game theory problems—but they also retain a more authentic meaning to actors, one that comports with the perspective of most lay individuals.

Metapreferences as a conceptual tool in understanding free will

The concept of "metapreferences" becomes useful at this point in conceptualizing situations in which humans act freely, but (as noted above) in accordance with certain logically understandable goals. By this name or others (such as meta-rankings; or first-, second-, third-, etc., order preferences) metapreferences have been used in similar roles since at least Frankfurt (1971), cropping up prominently in Sen (1977), and are used extensively in Hollis (1983).[38] Frankfurt explains:

> [O]ne essential difference between persons and other creatures is to be found in the structure of a person's will. Human beings are not alone in having desires and motives, or in making choices. … It seems to be peculiarly characteristic of humans, however, that they are able to form what I shall call "second-order desires" or "desires of the second order."
>
> Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are.
>
> … Someone has a desire of the second order either when he wants simply to have a certain desire or when he wants a certain desire to be his will. [6–7, 10]

Whether one conceptualizes metapreferences as "preferences over preferences" or instead layered and potentially contradictory preferences, the notion is entirely commonsensical and no doubt familiar to virtually anyone: we all prefer, at least at some biological level, to eat food that is tasty but not healthful to food that is healthful but not tasty.[39] And yet

[38] For related, see Grofman and Uhlaner (1985), Nida-Rümelin (1991), George (1984, 1993), Hirschman (1985), Dowell et al. (2007), Brennan (1993).

[39] The difference between these two conceptualizations seems to be meaningless. Assuming a choice between two outcomes, the former view (metapreferences as preferences over preferences) indicates that one has a single preference over the choice between *A* and *B*. But, then, the individual has a single preference over the choice between *preference for A over B* and *preference for B over A*, and then potentially a preference for *preference for preference for A over B* and *preference for preference for B over*

anyone who cares about his or her health or appearance likely wishes, at least some of the time, that he or she preferred the opposite. Thus, in principle, all preferences may have a metapreference structure that goes to two or more levels of preferences.

As Frankfurt elaborates on, metapreferences are immediately useful in conceptualizing free will, because for any decision they underscore the argument that, in contrast to rational-choice models, a simple (first order) preference between options may not be enough to deduce the behavioral outcome. Will the dieter choose to defect and scarf down a cookie in secret or stay the course and snack on raw broccoli? Most of us have stood in that scenario, feeling internal conflict over conflicting preference and meta-preference, and consequently making a "torn decision," as Balaguer borrows from the vernacular. It isn't that one's utility from eating the cookie or utility from eating the broccoli (or, more accurately, from the health benefits of doing so) are instable; it is that, in contrast to the rational-choice implication of humans as mechanical throughputs executing instant utility-maximization decisions, there is a moment of choice in which we potentially *recognize, select, and instantiate* a revealed preference all by virtue of the action we take.

Of course, if, at the deepest level, our selection is predetermined anyway, then (as was discussed in the preceding section) whether we describe such choice in the terms of

---

A. In other words, rather than treating metapreferences as a series of increasingly abstract preferences over the same outcome, this view treats metapreferences as preferences over increasingly abstract outcomes. In the second conceptualization, metapreferences as preferences of increasingly higher order, the opposite is the case: one may prefer *A* over *B* in the first order, *B* over *A* in the second order, *A* over *B* in the third order, and so on.

Homo economicus or *Homo psychologicus*, we are merely dancing around the actual causal story taking place outside our cognition (no matter whether it influences our cognition).

If, however, humans possess free will—as the rest of this paper assumes—metapreferences become an immensely useful way to model the decisional context surrounding the exertion of free will. This is because, as was stated as this paper's assumption nine: "possessing free will would [not] mean that humans cannot still undertake subconscious, habitual, reflexive, and otherwise unfree/predetermined actions. … [I]n nearly any scenario, an individual human may have a materially/biologically/rationally/etc.-given "default" behavior." The proposition that indeterminacy lies at the heart of human action does not free humans from the pressures of material forces—nor any more of those purely ideational; likewise, neither from socially given compulsions nor from the impulses, rational or irrational, spoken in one's own psyche. Thus social scientists can take these as a starting point, not for predicting social outcomes nor even explaining, without problematization, their causal processes, but rather as the fodder that free wills actually consume in driving human action.

Moreover, in most cases, there are intertemporal aspects to metapreference structures as well. For example, in the "dieter's dilemma," the utility from eating a cookie is received now, while the utility from eating the more healthful broccoli is accrued in the form of long-term health benefits.[40] This point is not merely (or even primarily)

[40] In other words, from a modeling perspective, utility would be modeled as a time-sequence, such as an ordered pair of (utility today, utility tomorrow) for every choice, or else as a function describing the change in utility each choice produces as a function of time.

important in that it would allow using discount rates and related techniques to model the effects of time on utility; it is important because it introduces the effect that sequences of actions can have on our preferences, and our metapreferences, over time.

Addiction is the premier example of the interrelatedness of decisions in the context of time and metapreferences. As Frankfurt puts it, "The unwilling addict has conflicting first-order desires: he wants to take the drug, and he also wants to refrain from taking it" (12); Sen speaks for the addict, "Given my current tastes, I am better off with heroin, but having heroin leads me to addiction, and I would have preferred not to have these tastes" (339). On its surface, the single first decision between consuming an addictive drug and abstaining is not as significant, even when metapreferences are factored in, as it becomes in the context of how that decision *alters* both first- (especially) but also potentially higher-order preferences in the future. Though counterintuitive in a mathematical sense, repeated consumption of an addictive drug will likely increase both the addict's first-order preference *to* consume the drug and, simultaneously, the addict's second-order preference to *not* consume the drug. In essence, with each indulgence, the addict is causing her lower- and higher-order preferences to diverge farther the next time she is faced with a decision about whether or not to consume.

While (again) metapreferences have conceptual use even to those who accept determinism, their benefit as a modeling technique is of greater use to voluntarists because of the normative import. Thinking about metapreferences and other closely related, commonly used concepts (such as autonomy and self-control/regulation) raises the perennial question for progressively minded social scientists: how do we accurately

take stock of what society is like now and encourage individuals to change it to something we would all prefer? Later, this paper applies that question to the study of war, but first, it examines more closely how technology can facilitate progress in human affairs.

The role of technology

Considering metapreferences leads to an interesting area of research: how science and, specifically, technology (that is, applications of scientific knowledge) may be used to manipulate behaviors by "locking in" certain preferences. In the extreme, technology actually *constructs* the social systems of tomorrow, and therefore

For example, in the case of addiction discussed above, repeated consumption of at least some addictive substances actually results in physiological changes in the brain that, of course, are part of what constitutes addiction.[41] Thus the very act of making certain choices, i.e., "instantiating" certain preferences to continue the thread from above, "cascades" into presenting certain menus of preferences instead of others in the future. Hard-to-kick habits are the more commonplace, less virulent versions, but the principle remains the same: decisions to feed or starve an unwanted habit, so to speak, alter the preference structure of each successive encounter one has with the habit.

Technology offers to play an interesting role, then, by enabling us to make decisions that interact with the natural/material/physical incentives that may present themselves to us (largely) first-order preferences. A simple example is an alarm clock.

[41] See Hyman and Malenka (2001), Nash (1997), and Volkow et al. (2004).

The process of sleep is a biologically driven process, but that humans maintain some control over. For example, someone waking up too early has the capacity to either rise and start the day or stay in bed. The technology of the alarm clock introduces a new variable—noise—that can influence the relative preferences. The key is that this technology can be initiated *in advance* of the scenario in which it is needed; the alarm can be set when the individual knows she will *prefer* to be awakened early, even though in the actual moment, were it not for the alarm, she would probably prefer to sleep longer.

Thus while Waltz argues that "No human order is proof against violence" (103), this raises the question of whether he would permit technological change to alter this statement. (It already plays a minimal/nonexistent role in Waltz (1979).) Without explicitly constraining the capabilities of technology—a difficult task to be sure, given the seemingly infinite frontiers in scientific and technological research—neither Waltz nor anyone can say definitively how technology can modulate what humans, alone and in groups, can and can't do.

As noted above, even from the voluntarist viewpoint, society—via human individuals—is built on material foundations. Ideas (such as norms, cultures of anarchy, and the like) are communicated socially through our biologically given senses; deliberation occurs with neural pathways; and the objects of social discourse—e.g., resources, food, and even human bodies—retain physical existence immutable by ideas (even if we accept social constructivists' claims about the power of ideas). In a sense (one sometimes ignored by their practitioners), the truth of social constructivism and other ideational schools of thought in social science is determined on material grounds.

Most readers will have little troubling agreeing that *some* of the material foundations are changeable; streams can be dammed and rerouted, mountains and forests can be razed, and cures can be found for disease. It may be easy to see the inventions and technologies of the past as "obvious" while maintaining skepticism when hearing of the mental flights of futurists. But do we have no grounds for saying any specific technological development is not possible at some point in the future? Even re-engineering humans to be more peaceful or loving, for example, cannot be ruled out. Moreover, because of the somewhat stochastic nature of technological revolutions, we cannot say with certainty that a material constraint of today will not be manipulable tomorrow. At best, we can only speculate which aspects of the material world are closer to or farther from engineers' reach.

Furthermore, and in accordance with the voluntarist perspective, insofar as social outcomes (i.e., laws of social behavior, generalized findings of social scientists, etc.) are determinate (even probabilistically or indirectly), it is because of constraints imposed by the material foundations of society. No measure of free will can prevent an isolated society that has produced no food from experiencing famine.

Similarly, man-made technologies, by modifying (in a sense) the natural world and the reach of man's control over it, affect human society even by voluntarist standards. Consider, for example, the ongoing technological development of weapons systems. Advanced weapons development may seem unlikely to upend global politics, given that military technology is generally expensive and closely guarded. Nevertheless, Rosenberg and Birdzell (1986) trace the decline of feudalism (and, indirectly, the rise of

the state system) to the development of the siege cannon; more recently, Deudney (2000) lets geographical and technological variables "arbitrate" what international security arrangements occur (in a sort of souped-up version of offense/defense balance theory). And while some have labeled nuclear warheads (and the logic of MAD) the "ultimate weapon" that has helped perpetuate the state system, anti-ballistic missile systems are clearly seen by at least some states (e.g., Russia) as a viable (or viable enough) antidote to MAD.

But the potential of technology raises a major challenge to the problem of scarcity as well: the finitude of provision is not theoretically true *a priori*; it is an empirical fact of our world. If clean water, for example, can be produced at effectively no cost, then water should not produce conflict. The same goes for food, territory (which is usually desired because of its productive value, anyway), or any typical consumable good. Therefore any theory of human conflict that traces back to scarcity is assailable on this point.

Moreover, this is different from the claim that a harmony of interests exists, e.g., between trading states. Rather, there need not be a harmony of interests, but rather there only need be technology to fulfill disparate interests simultaneously. Such technology would challenge the idea that *any* good is zero-sum in nature. For example, consider prestige; assume all individuals want to see their state as a superpower, and clearly all states cannot simultaneously have superpower status in the current world. Or, consider the conflict over the Temple Mount site in Jerusalem, where multiple religious groups

desire exclusive control over a particular piece of land. Is it possible for all religious groups to simultaneously live in a world where their religious group has exclusive control of the Temple Mount?

For all of human history, our wants—whether satiable or insatiable—have been met (or not) based on the rules of the macroscopic "real" world. But it is increasingly plausible, and perhaps probable, that humans of the future will not live exclusively or even primarily in this world (or, more accurately, in a world ruled by the natural laws we have grown used to). Such virtual worlds—computer-generated, program-run simulations of the real world—have the power to undo the accuracy or the relevance (or both) of virtually any social theories. While computers are still far from truly immersive, "augmented reality" applications are becoming commonplace on smartphones and tablet computers, and scientists continue to make progress in allowing human subjects to control computer objects through direct thought (see, e.g., Hochberg LR et al. [2006]). Moreover, the equivalent real-world size of the economy of one popular virtual world, *Second Life*, already exceeds a number of independent states (e.g., Comoros, East Timor, Tonga), and its real-world equivalent land mass is roughly twice that of Hong Kong. While the latest rash of 3-D television and movies may indeed be a technological fad, such fads only obscure an underlying virtualization of society.

For now, this virtualization has had its greatest impact in the fields of entertainment, information provision, and communication. Further, the virtualization still relies on actors outside of the virtual worlds to give the virtual world meaning. Hence, while shutting down all the virtual banking of an enemy state would have some military

impact, this impact remains different from bombing a tank factory. Nevertheless, the more virtual banking (for instance) is the target of warring states, the more likely will the best weapons be viruses, computer programs, and so forth—all of which "live" in the virtual world, too. And if the economic basis of a state's power likewise becomes quasi-virtual—for example, superior technological infrastructure or operating systems—then the more the targets of war will be virtual.

When it comes to social science, virtual worlds also offer a glimpse at possibilities in action. For example, the rules governing the economy of *Second Life* are market-based, and the company behind *Second Life* maintains monetary functions (through a virtual central bank and a real-world currency exchange). Thus, the social rules—e.g., the causal and constitutive laws—of *Second Life* come down to the programmatic/coding decisions made by its programmers. Thus, the economic laws that *Second Life* obeys are, to at least some extent, determined by the programmers' intentions. And—interestingly—*Second Life* residents can choose the degree of exposure they want to combat, war, and the like. That is not quite the same as saying that a *Second Life* virtual world negates the findings of social science, but it still suggests that humans can create worlds wherein certain social laws do not apply, or are somehow manipulated. My speculation is that once true immersion hits—in the sense that by donning a virtual reality helmet, an individual's experience in a virtual world becomes hardly discernible from experiencing the real world—human participation in virtual worlds will explode. With that comes human interaction with other humans in a purely virtual environment,

one in which the social rules that prevail are those of the virtual world's programmers. (For more see Schroeder 1996, 2010.)

How, then, may we attempt to understand the "technological construction of international relations? The question is how individuals, as well as governments, interact with technology.

For now, I offer a suggestive example. First, intuitively, we might say that insofar as a given technology can be used at time *t* by humans to accomplish specific purposes that have their result in time $t + 1$, the social and material rules at time *t* will determine who gets to use technology and how. New technology does not drop to humanity from the sky (the likes of Mir and Skylab—and perhaps Coca-Cola bottles—excluded); rather, technology is usually the result of years of research and experimentation, often including governmental support. To be good social scientists, therefore, we must investigate how the social structure of the current world shapes technological development, potentially reproducing itself even as the material world changes.

For example, Waltz thinks little of technology's ability to alter the relative balance of power (let alone the relevance of the concept); he assumes either that the newest technologies will always be developed in existing powers, or that great powers will always be able to steal or replicate new technology. "In shaping the behavior of nations, the perennial forces of politics are more important than the new military technology," Waltz writes (173), saying later, "Gunpowder did not blur the distinction between the great powers … nor have nuclear weapons done so" (181). For Waltz, inequitable infusion of military technology does not change the international political

arena because it does not "change the economic bases of a nation's power." (This strikes me as incredible, for on the domestic level it suggests Detroit should never have seen competition from Palo Alto.) This might be called the hardcore political realist attitude toward technology: the powers that be will always see to it that the technologies developed and used will be developed and used to benefit *them*, or at least so as to not hurt them; power maintains itself. If this were the case, whoever had power at the dawn of human civilization would never lose power due to technology; absent exogenous shocks, even as technology changes over time, the same actors have consistent power over time.

Challenging this hardcore Waltzian view, however, that technology will always empower leaders (positive feedback) rather than favoring those out-of-favor[42] (negative feedback), is the question of how well the powers that be can understand all the consequences of any given technology. Although in practice there is the fact that leaders are usually not scientists or engineers (some communist states excepted), the greater theoretical issue is the unintended consequences of technology.[43] The future of technology does not seem to be perfectly predictable by anyone, a fact that seems increasingly pertinent as computers approach the point at which they can program themselves (as discussed above). This suggests that if all that the holders of power desire

[42] We might consider this Gilpin's (1984) view, although as far as I am aware no one has examined this disagreement on technology at the heart of Waltzian vs. Gilpinian realism. As far as I see, both seem to err; Waltz writes as though the powers-that-be will have no trouble making technology work for them, while Gilpin writes as though the power-that-be (i.e., the hegemon) is wholly incapable of devising a system that prevents technology from working against the hegemon's superiority. Both seem to have empirical examples that work for their view.

[43] These unintended consequences are different than those explored by Pierce (2004), which emphasizes the way the powers-that-be can use the unintended consequences of technology *against* their enemies.

is power (or relative benefits), the best strategy is to stifle technological research so as to reduce the probability as low as possible that some technology will be developed that favors those out of power.[44] On the other hand, if the holders of power care about economic benefits, for example, or absolute gains/benefits, they may allow some level of technological research, even while patrolling it to reduce direct threats to their regime (with "regime" broadly construed).

Thus, clearly a regime such as North Korea's can more easily suppress or control technological developments, while a free-market-based liberal democracy will be faced more frequently with unexpected consequences from technology (though democracies have so far prevented almost any research into human cloning).[45] At one extreme, the most tyrannical state can easily suppress technological development *ad infinitum* and thereby perpetuate a single power structure. At the other extreme, an open society with a vibrant and well-financed R&D sector (either through benign government largesse or corporate investment) will lead to broad and sometimes unexpected technological developments, many with the potential to undermine some existing centers of authority. But even a global society allowing a moderate level of technological development seems unlikely to be able to perpetuate itself forever, given that technology, though a human creation, usually leads to unforeseen uses and consequences. More formally, the

[44] This is technically incorrect; more accurately, a tyrant might consider the possibility than any given technological research program will increase threats to the regime versus decreasing threats (i.e., allowing better monitoring/suppression of dissidents), then restrict only the technological research that seems more likely to hurt the regime than benefit it. Also, we might assume that even in a tyrannical regime, some secret technological research will continue, as well as incidental research that necessarily occurs through the frequent trial-and-error process of using existing technology and keeping it maintained.

[45] See Mackenzie (2011) for a recent glimpse into how liberal regimes deal with technology.

stochastic element in any technology's effects may be enough to disturb the distribution of power, especially across development cycles.

Interestingly, this same form of approach might be applied to the relationship between human nature and technology. If human nature can be changed by technology, then we shouldn't look in the rear-view mirror of human behavior to understand human nature *unless* there is something inherent about human nature that will cause us to inevitably redesign human nature in ways that leave it consistent with existing human nature (and that's assuming those redesigns do not spur their own unintended consequences). Thus, even if technology offers a sea of potential for humans to change our world and ourselves, the material world we inherited from pre-human days may be such that we only ever swim in the same old lagoon of intractable human nature.

Of course, I am not attempting to close the door on social processes that are working alongside underlying material/technological processes. For example, just because North Korean leaders are attempting to suppress technology in their own state, that does not imply those leaders are somehow isolated from normal IR processes (whether realists, liberals, constructivists, or critical theorists are most correct in describing those). Thus, if Wendt's world-state argument is correct, then the forces he sees working to eventually produce a world state may well have an independent effect on developed technology. Nevertheless, technology has an effect on the world-state reasoning; hence, there is a dualistic process whereby material (and technology) constructs the social world, and that social world then constructs (to a lesser degree,

perhaps) technology and thereby alters society's material foundations. (In a sense, this might be seen as a technological spin on Giddens' (1984) structuration theory.[46])

Technology, agency, ideas, and norms

The only way to predict the future is to create it, some say, and along these lines the increasing role of technology in society raises questions for social scientists even deeper than those discussed above.[47] After all, if technology gives humankind the ability to alter the determinants of social behavior, then technology may be seen a sort of "multiplier" of human agency.[48] Technology gives humans a freedom to define and modify our existence, both as individuals and as societies, unlike ever before.[49]

The potential of technology to change the material, social, economic, security, etc., landscapes, and the question of how or how well human actors can control and direct this potential, constitutes my focus in the remainder of this paper. If humans can control and direct technology, then in the hypothetical world in which all technologies are known and usable, humans can essentially choose to live in whatever social world they desire,

---

[46] That is, imagine a cycle between two things: society/social structure and the material foundations of society. The question is, to what degree can society change its foundations in a manner independent of how society is materially constructed? This is akin to the relationship of states and IOs; do the latter actually accomplish anything independently in their dealings with states if they themselves are created by states? What is interesting, I submit, is that many of social constructivists' ideas about the relationship between the individual and the (social) whole are closely related to the relationship between the social and the material, and in other work I hope to explore this connection by encouraging social scientists to think more about free will and determinism. Nevertheless, the key point here is that when thinking about the future, we can neither hold technological change constant and look at society, nor can we hold societal change constant and look at technology, since the social world is both slave to the material world and, through technology, master of it.

[47] Variations on this quote are attributed to Peter Drucker, Eric Hoffer, and Alan Kay.

[48] There is a quite different interpretation of the relationship between technology and agency offered in Schroeder (2007).

[49] Of course, this is only true if our discovery/invention and use of technology is, in itself, an act of agency rather than structurally determined; if we lack free will in our dealings with technology, then how can technology bestow novel agency upon us?

and all that matters to social scientists is knowing why humans would choose to make any instantiate any social world as true over any others. The human mission in this world is, essentially, to come up with well formed, coherent theories (good "stories" or models, in a sense), then use whatever technologies, programs, etc., to implement and sustain that story.

As has been argued so far in this paper, it is difficult to see how the empirical accuracy of any theory—mid-level or grand—of international relations (or social science) could be safely generalizable into the future given the potential of technological change. I should briefly head off a potential criticism that would be based on a misunderstanding of my theory. A critic might argue that there are certain logically coherent theories that have stood the test of time in international relations, and that even though we may not be able to predict beforehand just *how*, empirically, the theories will apply, once we understand the application, we can make predictions confidently. For example, Waltz's balancing propositions may or may not apply to the modern conception of states, the critic would admit, but nevertheless when we find entities competing under anarchy, some form of balancing occurs. Or consider offense-defense balance theory: perhaps we do not know whether tomorrow's weapons systems will confer an advantage to defense or offense, but we know that weaponry providing an offensive advantage will result in a more threatening international security environment than will weaponry providing a defensive advantage.

The rebuttal takes two steps. First, note that any truth claim divorced entirely from empirical facts becomes—at best—a sort of raw, meaningless truth claim stated in

abstract logic. Therefore to try to conceive of an "empirically free" or "purely theoretical" version of Waltzian balancing simply by omitting states as the balancing actors does not actually remove all empirical content from the theory; unpacking the meanings of balancing, anarchy, and the like necessitates the explication of a range of empirical phenomena, such as power, governance, and the like. *Truly* separating all that is empirical from something like balance-of-power theories, then, leaves us with nothing more than a handful of bound variables and logical connectives. Such theories must be underdetermined, in the sense that the existence of a system of variables implying a specific causal relationship between the variables cannot be evaluated without reference to empirical data.[50] In the twilight between a theory with empirical content and a theory without any, we can admit only the most vague connections between logical atoms and empirical concepts, and herein lies the second step of the rebuttal. When we look for causal relationships in social science, we understand that there is always a seemingly infinite causal chain occurring between any two variables, so that "anarchy causes balancing" really means something closer to "anarchy causes self-help causes uncertainty causes dilemmas causes leaders to be rewarded for cautious, defensive foreign policies causes leaders to be more worried about more powerful states causes balancing," and even within each causal connection in the latter fragment, we could unpack and critique the alleged causal chain; for example, Wendt (1992) could be seen as critiquing the causal chain between the first two terms, anarchy and self-help. While Wendt's critique originates from constructivism, we might imagine a parallel critique from a "material"

[50] This result may be seen as an interpretation of the Duhem–Quine thesis for security studies.

perspective (again, Deudney [2000] offers perhaps the best example), because if any causal relationship in social science must be "relating," in part, through material, and we cannot assume these causal pathways are impervious to technological alteration. Therefore even if we could conceive of the most abstract, least empirical version of balancing theory where A actually can refer to anarchy and B to balancing, it would be impossible to establish A ⇒ B without reference to fairly thick empirical truths about the purported causal pathway between anarchy and balancing.

Thus, any theory or model of the world that is well-formed and does not violate the most basic laws of logic, ultimately, is true with respect to some empirical "circumstance," we might call it, on earth. If we can conceive of how anarchy would lead to bandwagoning, for example, then we can imagine an empirical world (perhaps that in Schweller [1994]) in which anarchy does not lead so easily to balancing. Even seemingly absurd theoretical conclusions—e.g., that a defensive advantage could augment the security dilemma—are true in some plausible empirical circumstance. Granted, some conclusions are relatively more distant from today's empirical circumstance; for example, those that would require humans to act fundamentally different than we do today will all seem more unreasonable.

Nevertheless, the result is that when considering the ways in which technology both intentionally and unintentionally alters our empirical circumstances, what we actually are considering is the way in which changes in technology, and especially the use of technology, actually causes humankind to run through worlds of different theoretical truth, as if scanning through a microfiche. For example, suppose we have a

technology that allows us to alter what C "is," technologically speaking, in such a way that therefore determines the truth or falsehood of two statements: $A \Rightarrow C$ and $C \Rightarrow B$. Thus, the causal relationship $A \Rightarrow B$ may be true or false depending on how we use technology. Of course, this extremely simple example can be expounded upon greatly; with hundreds of variables and a modal logic, the greater ability our technology has to alter fundamental relationships, the greater mess technology makes of any existing social science findings.

The straightforward consequence of this is simply uncertainty, along the lines of Dequech (2004). His interpretation might be characterized as a "soft" form of uncertainty: for example, certain types of technology will lead to certain cognitive heuristics; technological changes will cause an adjustment period as one technological paradigm is unrooted and replaced by another; and so forth. In Dequech's formulation, however, people are still people, so to speak; the social, economic, and material landscape—against which technology appears and changes, leaving uncertainty in its wake—does not, itself, change (or, at least, Dequech sets this aside). The harder version of uncertainty I am presenting here focuses on the ways in which the underlying landscape is not only not impervious, but is actually increasingly malleable by each technological revolution. As I see it, this technological angle is the most relevant aspect of the many things Schweller (2010) sees as driving "ennui," although as I see it technological change can both increase and decrease what he labels "entropy" in international relations, even if at present we accept the argument that technology has increased entropy in IR.

If my interpretation is correct, and if humanity has even some small independent agency in its discovery and use of technology, it necessitates a normative turn for social scientists, not only because the positivist's *what is* may be entirely indeterminate in a future saturated with technology, but because the normative *what should be* should weigh heavily on the technologist whose new designs have the ability to shift us from one social world to another.[51] Like engineers, we should become less interested in understanding the particular social world around us today for its own sake (though historians will retain this duty) and more interested in inventing a particular social world for tomorrow. (Compare this argument to that found in Price [2008].)

This is the leading edge of the final conclusion: the day intelligent beings have discovered all technology is the day they have the ability to build any material world they desire; and if social/ideational worlds are or can be[52] direct or indirect consequences of the material world, these intelligent beings can build any social world they desire.[53]

Ultimately, my suspicion is that computer scientists may, to some extent, be giving us a more accurate conception of the future of the human world than social

---

[51] This is separate from the legitimate normative debate over the use of technology—the means rather than the ends, that is.

[52] The distinction here between "are" and "can be" may seem important, but I do not believe it is. Suppose some—but only some—social/ideational worlds are the inevitable consequence of a given material world. If the intelligent beings want to create one of these worlds, they can simply instantiate the material world from which it naturally follows. But what if the intelligent beings want to create one of the social/ideational worlds that does not flow directly from a given material world? Presumably they could somehow "prime" the actors in that world with beliefs, religion, etc., that would result in the right social/ideational world with some high probability. (However the more I think about this, the more I believe that all social/ideational worlds would be the direct result of some instantiable material world, temporarily setting aside the issue of free will.)

[53] I used the qualifier "at least, *ex nihilo*" above, and without spending too much time on this point I want to note that using technology to create a new material world or environment may be something different than using technology to change/reform an existing world. Similarly, this assumes the intelligent beings are of one mind, or that a single intelligent being can undertake the task itself. If multiple beings disagree on the use of technology, the end result will be a consequence of the social laws regulating disagreement (etc.) as generated by the material facts of the world those beings inhabit.

scientists are, because in the face of technology only a field like computer science focuses on the broad domain of what is logically possible in the limit (and, again, this is even more true when considering self-programming computers). Further, technological development will change the worlds within which technological development (and the rules thereof) occur.

The Waltzian discussed above constitutes one of the middle scenarios. If technology allows the powers-that-be to accrue power; where the powers-that-be are the people, technology will strengthen democracy; where the powers-that-be are despots, technology will strengthen despots. In the Gilpinian scenario, however, technology perpetually serves to level the playing field, at least partially: technological change can facilitate great power cycles and the existence of power transitions, or, if it enables individuals more than states, it can facilitate the entropic conclusions in Schweller (2010). Still, in these scenarios, changes in technology do not actually change social behavior; we merely see varying expressions of the same fundamental social laws across changing technology.

Two other middle scenarios are, first, one in which technology actually *becomes* the government (i.e., a "technocracy" of intelligent philosopher-king machines) and, second, a random-walk model where technological advances confer no *consistent* change in society, but rather offer glimpses of all the preceding possibilities. (Perhaps this is the situation we find ourselves in.) The latter may, in fact, look quite similar to Schweller's *ennui*.

Chapter 4: Free Will at War: Examples from International Conflict

Of course he has a knife, he always has a knife, we all have knives! It's 1183 and we're barbarians! How clear we make it. Oh, my piglets, we are the origins of war: not history's forces, nor the times, nor justice, nor the lack of it, nor causes, nor religions, nor ideas, nor kinds of government, nor any other thing. We are the killers. We breed wars. We carry it like syphilis inside. Dead bodies rot in field and stream because the living ones are rotten. For the love of God, can't we love one another just a little?—that's how peace begins. We have so much to love each other for. We have such possibilities, my children. We could change the world.

—spoken by Eleanor of Aquitaine in *The Lion in Winter*

Human conflict, and specifically interstate war, provides a backdrop to drive this argument further. "War, to be abolished, must be understood," says Deutsch (1970). In this section, I springboard off Fearon's (1995) seminal work about rationalist explanations for war to look at how war can be thought of as similar to an addiction using the metapreference modeling technique, and hence how technology and other related solutions may alleviate the problem.

Rationalist explanations for war?

Wars are costly, yet regrettably they recur. Fearon (1995) calls this the "central puzzle" of war, and inspired researchers to look beyond structural circumstances (e.g., power transitions) and necessary causes (e.g., anarchy) to better understand the causal mechanisms by which states cannot agree to a mutually preferable (at least, weakly), peaceful compromise on the eve of a self-evident, costly war. Specifically, why are two rational states (that, by definition, recognize the *ex post* inefficiency of war) unable to look down the game tree of a crisis and implement a resolution of their conflict that leaves both better off (again, at least weekly so) and avoids the manifold costs of war. In other words, why don't states achieve the Pareto-superior outcome?

"[O]ne can argue that even rational leaders who consider the risks and costs of war may end up fighting nonetheless," postulates Fearon (379) near the beginning of his influential article on rationalist explanations for war. He goes on to say:

> The common flaw of the standard rationalist arguments is that they fail either to address or to explain adequately what prevents leaders from reaching *ex ante* (prewar) bargains that would avoid the costs and risks of fighting. A coherent rationalist explanation for war must do more than give reasons why armed conflict might appear an attractive option to a rational leader under some circumstances—it must show why states are unable to locate an alternative outcome that both would prefer to a fight. [380]

After arguing that five standardly given (at the time) "rationalist" explanations for war did not meet his coherency standard, Fearon proposes three new explanations.[54]

[54] Two of these three are directly connected, but more coherently stated, versions of two of the five original explanations considered (383).

Drawing on the preceding, and especially the discussion of metapreferences, time-inconsistent preferences, and addiction as well as technology, this subsection critiques Fearon's treatment of the central puzzle as remaining, in a dual sense, "nearsighted." First, in the more literal sense, Fearon draws researchers' focus to the immediate threat of war as an "outside option" to crisis bargaining, thereby framing war primarily as bargaining failure. Second, in the figurative sense of myopia, this interpretation implies that to avoid war, rational actors should focus on avoiding the breakdown of political bargaining.

To summarize the below: this paper argues that the nearsighted interpretation of the central puzzle is overshadowed by a parallel, and certainly no less riddling question: why are rational actors unable to make binding agreements not to make war with one another *before* crises occurs? In other words, if states know that war is inefficient, why can they not take war off the table as an option (whether outside or inside, as other models have it) to bargaining before a breakdown can occur?

Theoretical background: A (very short) history of violence

The proposed causes of war are far ranging, from deep, damning human flaws such as the Freudian "death drive" (Freud 1933) to the tragic consequences of misperception leading to "war nobody wants" (White 1968). That makes it necessary to at least briefly categorize causes of war and specify which this paper addresses.

Clearly, though, wars do occur, and therefore there can only be two explanations given the assumption of rational, unitary actors. First is the possibility that states may in

certain situations see war—destructive as it is—as nevertheless less costly than the result of a bargain that avoids war.[55] By assumption most rationalist IR scholars do not allow for this possibility, and I follow in adopting this assumption.

This leaves (second) the possibilities proposed by Fearon, and further explored by others: war occurs because of a breakdown in bargaining. This may occur for such reasons as the incentive to conceal private information (especially tactical advantages), disagreements about relative capabilities, the incentive to defend one's reputation, the inability to credibly commit to uphold a bargain, or the indivisibility of the object(s) in dispute. (The relative relevance of each reason is debated to some extent.) The post-Fearon paradigm, while departing from Cold War-era security studies, reinforces the earlier consensus that most wars are tragic; indeed, it strengthens that perspective by better illuminating the theoretical path to war.

Fearon offers a thoughtful tripartite categorization of war: "[O]ne can argue that people (and state leaders in particular) are sometimes or always irrational … [O]ne can argue that the leaders who order war enjoy its benefits but do not pay the costs … . [Or], one can argue that even rational leaders who consider the risks and costs of war may end up fighting nonetheless" (379).

First, we may note that Fearon ignores (deliberately) the possibility of *efficient war*. In this case, *contra* the assumptions of rationalist researchers, war may be an efficient outcome for societies. Because of the destructiveness of modern war, this efficiency likely does not simply refer to the stimulation of economic production, but

[55] Closely related is war between risk-acceptant states, in which case the gamble of war may be preferred—even with costs factored in—to a riskless bargain.

must refer to intangible benefits (or their tangible consequences) such as in-group cohesion, patriotism, the fulfillment of moral obligation, etc. Some of these fall under the heading of war between rational, but nevertheless risk-acceptant states, who may all prefer taking the gamble of costly war over the status quo. Some of these (e.g., a zero-sumness in the fulfillment of moral obligations) are parallel to the "divisibility" explanation for war between rationalists that Fearon offers, while others (e.g., risk-acceptant states) match, in part, Fearon's treatment of war due to private information (specifically, the possibility of mutual optimism, explained further by Wagner [2000] and Smith and Stam [2004]).

Further, by relaxing the unitary actor assumption, this category covers situations in which those who decide to make war can not be made to internalize the costs of war; e.g., war between autocrats (as Kant lamented) and diversionary war (although the "rally-round-the-flag" effect may be strong enough to fall under the "patriotism" explanation above). This thereby encompasses the second grouping Fearon names.

Importantly, across these explanations, the value of war to one party must be greater than the value of avoiding war to the other party, or else something must impede the possibility of side payments.

Two interesting points appear given this categorization. First, modeling attitudes toward risk is highly relevant to conflict studies because in nearly all the causes of war above, risk attitudes may play a role. Yet a certain attitude toward risk does not always translate into a greater likelihood of war or peace. Risk acceptance may make war efficient *ex ante* for all parties if, for example, states are fighting over the implementation

of indivisible (or, at least, believed to be so) moral principles. By contrast, risk acceptance could reduce the probability of accidental war in a preventative war situation. Therefore, although these are informal intuitions, all purported causes of war must be carefully examined for their sensitivity to presumed risk attitudes.

Also, nearly all situations of conflict imply that states consume utils over some span of time: for example, during hostilities, and then after war's resolution in the postwar peace; in rally-round-the-flag models, during the initiation of war and then after casualty reports roll in (which may be seen as a compatible, but somewhat heterodox, interpretation of Baum and Groeling [2010]). This suggests models of war must better account for individuals' discount rates on future utility. After all, few dispute *ex post* that war is costly; the question is to what extent the present value of war on its eve incorporates future expected costs.[56]

Of course, what may seem obvious but is nonetheless important in the context of this paper is that the fundamental problem of *ex post* inefficient war is that humans can not, universally anyway, prevent themselves from engaging in it. That is, the material and ideational construction of the world make it effectively impossible to forswear violence; credible commitment not to harm is currently bounded by this rather Hobbesian point.[57]

[56] This even suggests what would likely be a controversial assumption to economists: war is both efficient and inefficient, depending on which temporal side of it one is on.

[57] This implies that in almost all reasonable circumstances, the possibility of efficient war will always exist, if only because humans can always invent ideological frameworks that create zero-sum, indivisible disputes. Interestingly, this seems the most potent long-term source of conflict and war, despite Fearon brushing it aside due to the possibility of side payments. Also note that this remains possible even if Hobbesian and Freudian explanations of human conflict are wrong, though there may be some escape through conclusions about epistemology (i.e., to what extent human belief is bounded or determinate).

Interpretations of the central puzzle

While acknowledging not only the work that Fearon and company, but also that many other conflict theorists, accomplished by "zooming in" on the theoretical onset of war and pointing out the incentives to reach a peaceful bargain, this paper suggests we should "zoom out" again and focus on why those same states have not reached a solution to take the option of war off the table for good.[58] As asked previously, why does the nearsighted interpretation of the central puzzle overshadow the farsighted interpretation?

Two common criticisms of this project may be that this paper ignores selection effects or has omitted that the same "just-in-time" causes of war obstruct an ahead-of-time prevention of war. The former point suggests that insofar as rational states are able to arrive at mechanisms ahead of time to avoid war in the case of conflict, they have; the democratic peace or the promise of sanctions from IOs might be named examples. Nevertheless, claiming that those dyads able to arrive at agreements to avoid war already have (and that those that haven't cannot) does not, by itself, explain how; and if it does explain how, the question is why other dyads cannot emulate the same mechanisms in order to avoid inefficient war.

As to the latter point, it is true that states may not be able to agree ahead of time to bind themselves from making war for reasons similar to those that push them over the brink into war in the first place. But when we zoom out, the same categories that prevent just-in-time peace agreements do not mirror themselves in preventing ahead-of time

[58] In addition to avoiding the prospect of inefficient war in the future, such a solution should presumably improve economic relations and reduce the chance that an interstate crisis, even if not leading to war, could scare markets with the threat of war.

peace agreements. If states or leaders recognize that war is Pareto inefficient and unavoidable in certain structural situations, then given a cost-free mechanism to avoid those situations, they should adopt it. In this sense, the decision to avoid war occurs before states know whether future wars will occur due to informational/accidental, divisibility, commitment, or whatever other reasons, or whether those reasons will be due to irrational or rational decision making. States only know that such land mines exist and that most represent Pareto inferior outcomes.

Given that states do not adopt a mechanism in certain circumstances, we can only accept that such a mechanism does not exist or that vis-à-vis certain other states, war must be efficient.[59] Given that such mechanisms exist on the domestic level, it must be either that war *is* efficient (as discussed above) or that something about the state system with the risk of war makes it preferable to a hierarchical system without the risk of war.

This hits squarely at one of Gilpin's central themes, as revisited by Ikenberry (1999, 2001) and Hurd (1999): the creation of a world system by a hegemon, which has sufficient power and prestige to dominate the world system in the wake of hegemonic war. Again, given the existence of a cost-free mechanism (as may exists in domestic politics[60]) to prevent future wars, why can the hegemon not implement the mechanism to

[59] Again, this point needs clarity. The intuition is that if states do not accept a cost-free mechanism to avoid war, then the expected utility from taking a long-term "war lottery" (in other words, a lottery that may result in peace or may result in many different war lotteries) is that there is a positive expected utility from the war lotteries. This seems reasonable to see in the case of a democratic hegemon given the preponderance of power coupled with moral crusading behavior (hence, issue indivisibility). This also fits with cases of risk-acceptant revisionist states.

[60] This is another point to be drawn out given the existence of civil wars.

lock-in the postwar status quo?[61] That is, why can't the hegemon initiate rules or take measures so that war never comes again and the hegemon is always predominant? This marries Emerson and Niou's (1994) point that the global structure is not exogenous to states, but rather created by them, with Gilpin's emphasis on the hegemon's predominance.

This question is interesting apart from the empirical frequency of such actors in any given era so long as such actors do exist from time to time and have an interest in security cooperation with one another. Therefore this project is somewhat incompatible with Mearsheimer's (2003) theoretical view. Putting it another way, what keeps any group of security-seeking, status-quo states from upholding any collective security agreement among themselves or imposing it on others insofar as they are able, transforming that agreement into a self-sustaining "status quo equilibrium" to maintain perpetual peace among themselves?[62] The goal is to permanently divorce interstate politics from the diplomacy of violence (and unlink war from bargaining, which per Wagner removes its primary justification); as long as violence remains linked to bargaining states have an incentive not only to fight, but to engage in arms races and the like.

---

[61] A closely related question, indeed, is why the hegemon cannot create a system that either avoids or offers it the benefits of the law of uneven growth among states.

[62] A natural question here is what will drive a bargaining outcome absent the threat of either limited or total war, to borrow from Von Clausewitz. The answer should be the same factors that drive most domestic compromises and some international bargains: economic incentives and side payments, votes, norms, popular support, social movements, and the like. Bargaining would be relatively free of military control and would be determined by economic and social/normative factors, as mediated through whatever political system existed when the status quo was locked-in (or through the political system created through the political system that had been locked in, or otherwise peaceably changed).

I should head off criticism from those who would argue that such a mechanism would be effective most when needed least, and vice versa. This misses the point: that if such a mechanism is adopted, even if when needed least—for example in what Kegley (1993) might call a "neoidealist moment"—then it should remain effective for when it is needed most.[63]

Metapreferences, technology, and war: war as curable addiction

Consider Fearon's argument that

> The conventional distinction between wanted and unwanted wars misunderstands the puzzle posed by war. The reason is that the standard conception does not distinguish between two types of efficiency—*ex ante* and *ex post*. As long as both sides suffer some costs for fighting, then war is always inefficient *ex post*[.] [383]

The distinction between *ex ante* and *ex post* in terms of the costs of war is indeed important—but it goes beyond Fearon's treatment. It makes a difference whether, even if one recognizes a cost associated with the consumption of a good, that cost is temporally separated from the benefit; this is precisely the situation encountered with the dieter's dilemma discussed above, and more malignantly with addiction. Addictive behavior, though almost universally acknowledged (indeed, often even by the addict) as destructive overall, results in chemical stimulation of the brain's pleasure centers. The addict encounters—again, temporally separated—a strong high followed by the crash; i.e., in

[63] Even in the absence of economic interdependence or hegemonic domination, such a moment may occur when all states are at roughly equal levels of economic development with similar "steady state" growth rates, in which case no state can expect to naturally outgrow the military capabilities of another state. Of course, this assumes a Waltzian world in which economic capabilities are the ultimate determinant of military capabilities; see Waltz (1979).

economic terms, the addict consumes significant positive utility at time t_1 but is bound to consume significant negative utility at time t_2. (The decision can be considered to have happened at time t_0.) That the net utility over all time periods is negative does not on its own—automatically or mechanistically, anyway—help most addicts escape the clutch of addiction; but neither is addiction necessarily irrational.[64]

This is, or at least intuitively appears similar to, the problem states face with war: a divergent *ex ante* and *ex post* efficiency (net utility), with important similarities implicit in Fearon's central puzzle that (1) by definition, rational states recognize this scenario, and that (2) empirically, states have not been able to avoid engaging in inefficient war time and time again. The puzzle, then, when zooming out from the actual failure of bargaining on the eve of war, is *how do states overcome what amounts to an addiction to war?*[65]

Moreover, this complexifies the question of whether states *want* war. States may well "want" war in the same way that someone legitimately wants a night of carousing even though this "want" is strongly regretted the next morning. The literature that Fearon critiques, which argues that "many or perhaps most wars were simply wanted" (383), is only part of the picture, but so is the purely rationalist perspective that inefficient wars are all unwanted because they result, *ex post*, in deadweight loss. To understand war—the human compulsion for it, even in a procedural-rationality context, as well as the great regret of it—we must accept both sides of the coin.

---

[64] For discussions of "rational addiction," see, e.g., Stigler and Becker (1977) and Becker and Murphy (1988).

[65] Although discussed earlier, this does rely on the assumptions that states are rational and that war (or, at least, some of the wars used as the empirical evidence in point [2]), is inefficient. Explanations using different assumptions may well be correct, but they are outside the intended scope of this section.

To be clear, this is not to say that states "enjoy the activity of fighting for its own sake," or at least not necessarily. Even before war, states may recognize the downside to fighting. What is important is the change in the utility for war: from either positive or weakly negative to much more significantly negative. (And as Fearon notes, even if both states recognize the inefficiency of war, the higher the costs, the easier it is to locate Pareto-optimal bargains. Therefore even with states recognizing the cost of war at the outset, the bargaining range can be reduced. After a conflict, even the victorious state, recognizing and paying the costs of war, may look back and wish it had pushed more firmly for a bargained outcome.)

Additionally, this scenario is furthered by most of the mechanisms Fearon offers to explain war, because most can be modeled as a prisoner's dilemma.

Chapter 5: Conclusion

The incompatibility between free will and determinism, and the relative ignorance of either perspective by international relations scholars, has significant implications for how we study international relations. In this paper I have attempted to bring this argument to fruition by explaining how and showing why it matters, as well as offering speculative thoughts about how international relations and other political science scholarship forge onward in new directions based on what we can learn from the behavioral determinacy debate.

As mentioned near the beginning of this paper, most of this argument has been predicated on the assumption that free will and determinism are mutually exclusive possibilities (i.e., the incompatibilist position), and some of my points are built upon the argument that we do have free will (i.e., libertarian position). There is a great backdrop of millennia of thinking on these topics, and of course this paper only glosses.

Many readers may bemoan my work as simply proffering up more "wishful thinking" without recognizing its dangers. Insofar as I have not closed the door entirely on determinism, I understand the fear, best expressed by Carr, of naive utopianism and unwarranted optimism. At the same time, there is the reverse risk of "naive realism" and unwarranted pessimism. Which is worse: wishful thinking or never wishing at all?

References

Abouzeid R (2011) Bouazizi: the man who set himself and Tunisia on fire. *TIME Magazine*. Retrieved July 12, 2011, from http://www.time.com/time/magazine/article/0,9171,2044723,00.html.

Achen CH (2002) Toward a new political methodology: microfoundations and ART. *Annu R Polit Sci* 5: 423–50.

Adcock R (2007) Who's afraid of determinism? The ambivalence of macro-historical inquiry. *J Phil Hist* 1: 346–64.

Anon. (2010) Witnesses report rioting in Tunisian town. Reuters. Retrieved July 14, 2011, from http://af.reuters.com/article/topNews/idAFJOE6BI06U20101219.

Anon. (2011) Mohamed Bouazizi: memories of a Tunisian martyr. BBC News. Retrieved July 12, 2011, from http://www.bbc.co.uk/news/world-africa-12241082.

Archer MS (1982) Morphogenesis versus structuration: on combining structure and action. *Brit J Sociol* 33: 455–83.

Atiq EH (2012) How folk beliefs about free will influence sentencing: a new target for the neuro-determinist critics of criminal law. *New Crim Law Rev* 16:449–93.

Axelrod R (1981) The emergence of cooperation among egoists. *Amer Polit Sci R* 75: 306–18.

Böök L (1999) Toward a theory of reflexive intentional systems. *Synthese* 118: 105–17.

Balaguer M (2010) *Free Will as an Open Scientific Problem*. Cambridge: MIT Press.

Bandura A (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, N.J.: Prentice-Hall.

Baum MA Groeling T (2010) Reality asserts itself: public opinion on Iraq and the elasticity of reality. *Int Org* 64: 443–79.

Becker GS Murphy KM (1988) A theory of rational addiction. *J Polit Econ* 96:675–700.

Berger P Pullberg S (1965) Reification and the sociological critique of consciousness. *Hist Theor* 4:196–211.

Bigo D (2013) International political sociology. In Williams PD, ed., *Security Studies: An Introduction*. New York: Routledge.

Bueno de Mesquita B (2009) *The Predictioneer's Game.* New York: Random House.

Burns K Bechara A (2007) Decision making and free will: a neuroscience perspective. *Behav Sci Law* 25: 263–280.

Byrne E (2011) Death of a street seller that set off an uprising. *Financial Times*. Retrieved July 12, 2011, from http://www.ft.com/cms/s/0/6ed028a2-21a2-11e0-9e3b-00144feab49a.html.

Carr EH (1939) *The Twenty Years' Crisis, 1919–1939*. London: MacMillan.

Cartwright N (2009) If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis* 70: 45–58.

Cohen GA (2000) *Karl Marx's Theory of History: A Defence*. Princeton: Princeton University Press.

Cotton M (2005) A foolish consistency: keeping determinism out of the criminal law. *Bost Univ Pub Int Law J* 15: 1–48.

Davies P (2004) Undermining free will. *Foreign Policy* 144: 36–38.

Dennett DC (2003) *Freedom Evolves*. New York: Penguin.

Dequech D (2004) Uncertainty: individuals, institutions and technology. *Cambridge J Econ* 28: 365–78.

Deudney D (2000) Geopolitics as theory: historical security materialism. *Euro J Int Relat* 6(1): 77–107.

Deutsch K (1970) Quincy Wright's contribution to the study of war. *J Conflict Resol* 14:473–478.

Doty RL (1997) Aporia: a critical exploration of the agent–structure problematique in international relations theory. *Eur J Int Relat* 3: 365–92.

Fahim K (2011) Slap to a man's pride set off tumult in Tunisia. *New York Times*. Retrieved July 12, 2011, from http://www.nytimes.com/2011/01/22/world/africa/22sidi.html.

Fearon JD (1995) Rationalist explanations for war. *Int Organ* 49: 379–414.

Feaver PD et al (2000) Brother, can you spare a paradigm? (Or was anybody ever a realist?) *Int Sec* 25: 165–93.

Feinberg, Joel. 1986. *Harm to Self.* New York: Oxford UP.

Flyvbjerg B (2001) *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again*. Cambridge: Cambridge University Press.

——— (2004) A Perestroikan straw man answers back: David Laitin and phronetic political science. *Polit Soc* 32: 389–416.

Flyvbjerg B et al, ed. (2012) *Real Social Science: Applied Phronesis*. Cambridge: Cambridge University Press.

Frankfurt HG (1971) Freedom of the will and the concept of a person. *J Phil* 68:5–20.

Freud S, Einstein A (1933) Why war? In Strachey J, ed., *Standard Edition of the Complete Psychological Works XXII, 197–215*.

Friedman M (1953) *Essays in Positive Economics*. Chicago: U Chicago Press.

——— (1957) *A Theory of the Consumption Function*. Princeton: Princeton UP.

Gartzke E (1999) War is in the error term. *Int Org* 53(3): 567–87.

George D (1984) Meta-preferences: reconsidering contemporary notions of free choice. *Int J Soc Econ* 11(3/4):92–107.

George D (1993) Does the market create preferred preferences? *R Soc Econ* 51:323–46.

Giddens A (1984) *The Constitution of Society*. Berkeley: University of California Press.

Giplin R (1984) *War and Change in World Politics*. Cambridge: Cambridge University Press.

Glock CY (1964) Images of man and public opinion. *Pub Opin Q* 28:539–46.

Gould SJ (1988) Untitled quote in The meaning of life. *Life,* December 1988.

Grieco JM (1988) Anarchy and the limits of cooperation: a realist critique of the newest liberal institutionalism. *Int Organ* 42(3): 485–507.

Grofman B Uhlaner C (1985) Metapreferences and the reasons for stability in social choice: thoughts on broadening and clarifying the debate. *Theory Decision* 19: 31–50.

Hafner-Burton EM et al (2009) Network analysis for international relations. *Int Organ* 63: 559–92.

Hochberg LR et al. (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442: 164–71.

Hollis M (1983) Rational preferences. *Phil Forum* 14.

Hollis M Smith S (1990) *Explaining and Understanding International Relations*. Oxford: Clarendon Press.

Hopf T (1998) The promise of constructivism in international relations theory. *Int Sec* 23: 171–200.

Hopf T (2010) The logic of habit in International Relations. *Eur J Int Relat* 16: 539–61.

Hyman SE Malenka RC (2001) Addiction and the brain: the neurobiology of compulsion and its persistence. *Nature Reviews Neuroscience* 2:695–703.

Hurd I (1999) Legitimacy and authority in international politics. *Int Org* 53:379–408.

Ikenberry GJ (1999) Institutions, strategic restraint, and the persistence of American postwar order. *Int Sec* 23(3):43–78.

——— (2001) *After Victory: Institutions, Strategic Restraint, and the Rebuilding of Order after Major Wars*. Princeton: Princeton UP.

Inwagen PV (1983) *An Essay on Free Will*. Oxford: Clarendon Press.

Kegley CW (1993) The neoidealist moment in international studies? Realist myths and the new international realities. *Int Stud Q* 37(2): 131–46.

Kratochwil F (1989) *Rules, Norms, and Decisions: On the Conditions of Practical and Legal Reasoning in International Relations and Domestic Affairs*. Cambridge: Cambridge UP.

Kurki M (2006) Causes of a divided discipline: rethinking the concept of cause in International Relations theory. *R Int Stud* 32: 189–216.

Kurzman C (2004) Can understanding undermine explanation? The confused experience of revolution. *Phil Soc Sci* 34:328–351.

Lichbach MI (1996) *The Cooperator's Dilemma*. Ann Arbor: University of Michigan Press.

Lucas RE (1976) Econometric policy evaluation: a critique. In Brunner K Meltzer A, eds., *The Phillips Curve and Labor Markets*. Amsterdam: North-Holland Publishing.

Mackenzie I (2011) Sarkozy questions 'neutral' net at e-G8 forum. BBC News. Retrieved June 22, 2011, from http://www.bbc.co.uk/news/technology- 13518871.

McCarty NM Meirowitz A (2007) *Political Game Theory: An Introduction*. Cambridge: Cambridge UP.

McIntosh D (1995) *Self, Person, World*. Evanston: Northwestern University.

Mearsheimer JJ (2003) *The Tragedy of Great Power Politics*. New York: WW Norton.

——— (2005) E.H. Carr vs. Idealism: the battle rages on. *Int Relat* 19:139–52.

Merker B (2013) Freedom and normativity – varieties of free will. In Sellars WS Lehrer K, eds., *Autonomy and the Self*. Dordrecht: Springer.

Nash JM (1997) Why do people get hooked? Mounting evidence points to a powerful brain chemical called dopamine. *TIME*, May 5, 1997.

O'Hanlon S (2008) Towards a more reasonable approach to free will in criminal law. *Cardozo Pub Law Policy Ethics J* 7:395–407.

Price R (2008) Moral limit and possibility in world politics. *Int Organ* 62:191–220.

Rohr, MV (2011) The small Tunisian town that sparked the Arab revolution. *Der Spiegel*. Retrieved July 12, 2011, from http://www.spiegel.de/international/world/0,1518,751278,00.html.

Ryan Y (2011a) The tragic life of a street vendor. Aljazeera. Retrieved July 12, 1011, from http://english.aljazeera.net/indepth/features/2011/01/201111684242518839.html.

——— (2011b) How Tunisia's revolution began. Aljazeera. Retrieved July 14, 2011, from http://english.aljazeera.net/indepth/features/2011/01/201112612181598548 3.html.

Ryle G (1949). *The Concept of Mind*. London: Hutchinson & Company.

Sartre JP (1998) Existentialism and humanism. In Loptson P, ed., *Readings on Human Nature*. Peterborough, ON: Broadview Press.

Sartre JP (2000) *Jean-Paul Sartre: Basic Writings*. Priest S, ed. London: Routledge.

Sasso P (2009) Criminal responsibility in the age of "mind-reading." *Am Crim Law R* 46: 1191ff.

Schroeder R (1996) *Possible Worlds: The Social Dynamic of Virtual Reality Technology*. Boulder: Westview.

——— (2007) *Rethinking Science, Technology, and Social Change*. Stanford: Stanford University Press.

——— (2010) *Being There Together: Social Interaction in Virtual Environments*. Oxford: Oxford University Press.

Schweller RL (1994) Bandwagoning for profit: bringing the revisionist state back in. *Int Sec* 19(1): 72–107.

——— (2006) *Unanswered Threats*. Princeton: Princeton UP.

——— (2010) Ennui becomes us. *Natl Int* Jan./Feb. 2010.

Searle JR (1995) *The Construction of Social Reality*. New York: Free Press.

Seebaß G (2013) Freedom without choice? In Sellars WS Lehrer K, eds*., Autonomy and the Self*. Dordrecht: Springer.

Sen AK (1977) Rational fools: a critique of the behavioral foundations of economic theory. *Phil Pub Aff* 6: 317–44.

Skinner Q (1989) Language and political change. *Political innovation and Conceptual Change*. Cambridge: Cambridge UP.

Smith A Stam AC (2004) Bargaining and the nature of war. *J Conflict Resol* 48: 783–813.

Snyder GH (1997) *Alliance Politics*. Ithaca: Cornell University Press.

Stanley J Williamson T (2001) Knowing how. *J Phil*, 98:411–44.

Stigler GJ Becker GS (1977) De gustibus non est disputandum. *Amer Econ R* 67:76–90.

Tullock G (1981) Why so much stability? *Pub Choice* 37:189–205.

Viney W et al (1982) Attitudes toward punishment in relation to beliefs in free will and determinism. *Hum Relat* 35:939–49.

Vohs KD Schooler JW (2008). The value of believing in free will: encouraging a belief in determinism increases cheating. *Psychol Sci* 19:49–54.

Volkow ND Fowler JS Wang GJ (2004) The addicted human brain viewed in the light of imaging studies: brain circuits and treatment strategies. *Neuro Pharmacology* 47:3–13.

Wagner RH (2000) Bargaining and war. *Amer J Polit Sci* 44: 469–84.

Wallerstein I (1979) *The Capitalist World-Economy*. Cambridge: Cambridge UP.

——— (1997) Social science and the quest for a just society. *Amer J Sociol* 102: 1241–57.

Waltz KN (1954) *Man, the State, and War*. New York: Columbia UP.

——— (1979) *Theory of International Politics.* Boston: McGraw-Hill.

Weber M (1956), as translated in Winckelmann J, ed., *Gesammelte Aufsiitze zur Wissenschaftslehre*. Tübingen, Germ.: JCB Mohr.

Wendt AE (1987) The agent-structure problem in international relations theory. *Int Organ* 41:335–70.

——— (1991) Review: bridging the theory/meta-theory gap in international relations. *R Int Stud* 17:383–92.

——— (1992) Anarchy is what states make of it: the social construction of power politics. *Int Organ* 46391–425.

——— (1998) On constitution and causation in International Relations. *R Int Stud* 24: 101–18.

——— (1999) *Social Theory of International Politics*. Cambridge: Cambridge UP.

——— (2001) Driving with the rearview mirror: on the rational science of institutional design. *Int Org* 55:1019–49.

——— (2006). Social Theory as Cartesian science: an auto-critique from a quantum perspective. In Guzzini S Leander A eds., *Constructivism and International Relations: Alexander Wendt and his Critics*. Abingdon: Routledge.

Whitaker B (2010) How a man setting fire to himself sparked an uprising in Tunisia. *The (Manchester) Guardian*. Retrieved July 12, 2011, from http://www.guardian.co.uk/commentisfree/2010/dec/28/tunisia-ben-ali.

White RK (1968) *Nobody Wanted War: Misperception in Vietnam and Other Wars*. Garden City, NY: Doubleday.

Wight C (2013) Philosophy of social science and international relations. In Carlsnaes W Risse T Simmons BA, *Handbook of International Relations*. Los Angeles: SAGE.

Williams G Mayer J (1962) Determinism versus free will, and individual ethical subjectivism. *Social Science* 37:50–53.

Worth RF (2011) How a single match can ignite a revolution. *New York Times*. Retrieved July 12, 2011, from http://www.nytimes.com/2011/01/23/weekinreview/23worth.html.