CHAPTER NINE

Homo Atomicus: An Actor Worth Psychologizing?

The Problems of Applying Behavioral Economics to Nuclear Strategy

Anne I. Harrington and John Downer

The idea of the perfectly rational, utility-maximizing actor has done a lot of work for social scientists over the years. This is most visible in economics, where *homo economicus* has served as the hypothetical protagonist of innumerable models, theories, and formulae. Less visible but perhaps equally significant, however, has been the unnamed rational actor at the heart of modern nuclear strategy, invoked in the game-theoretic stratagems of Schelling[1] and the many influential theorists and policymakers who followed in his wake. Let us call this actor *homo atomicus.*

Homo atomicus enjoys considerable authority in the nuclear sphere, where rational actor models are still used to guide most strategy deliberations. This is perhaps surprising, however, since the authority of *homo economicus* has lately begun to wane[2] — in no small part due to the 2008 financial crisis, which dramatically undermined the rational actor models on which oversight of global financial system had been premised.[3]

The 2008 crisis, along with other perceived failures of foresight and leadership, raised important questions about rational actor theories in economics, and created space for a new model of economic man. Into it stepped the generation of behavioral economists such as Kahneman, Tversky, Thaler, Ariely, and Loewenstein. Drawing on a wide range of experiments in cognitive psychology, these thinkers exploded the notion of

homo economicus as a perfectly rational actor by highlighting persistent irrationalities in common economic behaviors.

One reason this new approach found a home in mainstream economics was that, even while it critiqued the rational protagonist of common economic models, it simultaneously offered ways to rescue the models themselves. Its key contribution was not to show that humans were often irrational, but to show that those irrationalities were shaped by consistent biases. This consistency allowed *homo economicus* to remain predictable even while he became irrational,[4] which was crucial because it meant that economists could tailor their models to anticipate his irrationalities. Behavioral economics didn't kill the rational actor, in other words, it simply gave him quirks.

If the models of economists can be improved by finessing their protagonist, however, then why not those of nuclear strategists? *Homo economicus* closely resembles *homo atomicus*, after all, and nobody imagines that the irrationalities identified by cognitive psychologists apply exclusively to the economic sphere. It seems intuitive, at least, that the models of nuclear strategists might evince the same shortcomings as those of economists, and, as such, that they might equally learn from cognitive psychology. Perhaps there is potential for a discipline of "behavioral deterrence," akin to behavioral economics.

This is the question that motivates this chapter. The goal is to evaluate the viability of cognitive and behavioral approaches to nuclear strategy.[5] Our conclusion strikes a skeptical note. Contra to many of the contributions to this volume, we will argue that cognitive psychology is markedly less useful to nuclear strategists than it is to economists. We do not doubt that nuclear strategic thought might gain from looking

beyond its (unquestionably problematic) commitment to rational actors, and from incorporating psychological considerations more directly in its analyses. We believe, however, that there are *a priori* reasons to imagine that insights into human irrationality have less purchase in the context of nuclear strategy than in economics.

The argument, in essence, will be that economists and nuclear strategists understand and justify their rational actor models in meaningfully different ways, with the effect that insights from cognitive psychology are more useful to one than the other. These differences in the ways that economists and strategists approach rational actors, and their significance, become visible if we unpack the specific meaning of "rational" and "actor" in each context. To this end, the crux of the chapter is divided into two parts:

Part 1 will look critically at the notion of rationality. It will begin by asking why it is, exactly, that economists and nuclear strategists felt comfortable assuming rational behavior for so long. It will then look at the specific circumstances under which it is logical for theorists to modify their rationality assumptions to accommodate cognitive biases. The conclusion of this part will be that it is only logical to adjust for cognitive biases under specific circumstances, and that these circumstances pertain to many economics problems but very few (if any) nuclear strategic problems. The upshot being that cognitive psychology has significantly less to offer strategists than it does economists, even if there remains scope to imagine it playing a role.

Part 2 will look critically at the notion of actors. It will examine who, or what, those actors represent in the relative contexts of economics and nuclear strategy. It will then examine the applicability of cognitive psychology to these actors at different levels of analysis. Its conclusion, similar to that above, will be that psychology has a more

complicated relationship to the actors that populate nuclear strategic models than it does to the models invoked by many economists. By drawing on studies that have examined similar issues, however, it will conclude by suggesting possible avenues for future research into cognitive psychology and nuclear strategy.

Part 1: On Rationality

An Unlikely Premise

In many schools of economics and nuclear strategy, rationality has become such a foundational premise that theorists often forget that it requires justification at all. This can be surprising to outsiders because, on some level, it seems intuitively obvious that people routinely behave in ways that would challenge even the most expansive definition of rationality. "All men are, at times, influenced by inexplicable sentiments," as the American novelist Charles Brockden Brown once put it.[6] Public health experts, for instance, have long struggled with the realization that most smokers already understand that their expensive habit is killing them.

Yet neither economists nor strategists are as blind to human nature as their critics sometimes suggest. Both have logical justifications for assuming rational actors despite the undeniable capriciousness of actual people.[7] Their argument, in essence, is not that all people act rationally, or even "boundedly rationally,"[8] all the time, but that theoretical models are able to transcend the messiness of real behavior. This argument tends to invoke at least one of four distinct justifications or mechanisms, each loosely corresponding to a specific construal of rationality and its purpose. For want of better labels we will call these: (i) "*Rationality as normative prescription*," (ii) "*Rationality as*

product of special circumstances," (iii) "*Rationality as product of systemic selection*," and (iv) "*Rationality as emergent mean or mode*." In what follows we will discuss these justifications in turn. In each case we will briefly outline the essential logic, its relationship to cognitive psychology, and its applicability to economic and strategic models.

i. Rationality as Normative Prescription

Some models invoke rational actors not to describe how actors *do* behave, but to explain how they *should* behave. Construed in this way, rational behavior is aspirational, and rational actor models are justified because they allow theorists to identify optimal solutions for problems that can then guide decision making.[9] In economics, for example, the game-theoretic Nash Equilibrium has been used to design auctions.[10]

This rationale is important to nuclear strategic models, many of which justify their invocation of rational actors in these terms, i.e. as a tool for exploring and guiding optimal decision making. Scholars routinely argue that deterrence theory served a normative function, for example.[11] From the earliest days of the Cold War, writes Morgan, "…[w]e needed instruction on how to do our best, to be rational, to avoid disaster but not lose."[12]

Crucially, however, cognitive psychology has little to offer models premised on this justification, and for a straightforward reason. Psychology can help researchers identify common biases, but in this context — where models are being used for guidance — the explicit purpose of formulating rational solutions is to *avoid* such biases.[13] Simply put: the fact that people tend to be poor at certain problems does not, in most instances, fundamentally change the optimal solution to those problems.

ii. Rationality as Product of Special Circumstances

A second justification for assuming rational actors conceives of models as being "descriptive" rather than "normative" (as in the justification above). It argues that even though actors often act irrationally, they nevertheless act rationally in special circumstances, such as when: a) they are incentivized to act rationally; b) the interpretation of what is rational is relatively clear and consistent to all; and c) they are provided resources — stratagems, information, advice, etc. — that promote and facilitate rational decision making.[14] Where objectives are clear, stakes are high, and actors are uncommonly judicious (or judiciously advised), in other words, then many theorists would argue that rationality becomes a more viable premise than it would be otherwise.[15] In economics, for example, the objectives of many financial decisions are relatively tightly defined, and where large sums are at stake and professional businesspeople involved, it arguably makes sense for economic models to assume that certain key decisions will be made on predictably rational grounds.

As with the previous rationale, the invocation of "special circumstances" offers a plausible justification for modeling nuclear strategic interactions in rational terms. This is because such interactions almost exclusively occur in circumstances where well-resourced actors are incentivized to make carefully considered decisions.[16] In questions of deterrence, for example, the high stakes of nuclear conflict undoubtedly create an incentive for states to reflect carefully on their decision making, defer to experts, and to formulate (and then follow) "rational" prescriptions. As Waltz succinctly puts it:

"…nobody but an idiot can fail to comprehend [the] destructive force [of nuclear weapons]. How can leaders miscalculate?"[17]

As with the previous rationale, however, the insights of cognitive psychology have significant limitations in this context, and for straightforward and intuitive reasons. Simply put, situations that incentivize deliberate reflection, by well-resourced actors, about problems with clear stakes, are not conducive to unthinking cognitive errors. The most fateful nuclear strategic decisions have unambiguous and carefully explored consequences, in other words, and nobody makes them without due consideration.

iii. Rationality as Product of Systemic Selection

A third justification for invoking rational actors in models of human behavior is to assume the existence of a selective process that actively favors rational outcomes.[18] Such processes are usually assumed to take the form of an external sorting mechanism — an "invisible hand," such as the free market — that advantages "rational" behaviors.[19] In economics, for instance, we might reasonably imagine that small businesses disproportionately fail when they make "irrational" economic decisions, and disproportionately prosper when they make "rational" economic decisions. And in these (eminently plausible) circumstances, it is logical to assume that economically "rational" businesses become more prevalent over time — boosted by the market's internal selection processes.[20] (Or, framed slightly differently, it would be reasonable to assume most established businesses to be rational actors, because if they were not rational then they would not be established businesses.)

While such assumptions are plausible in economic scenarios like that above, however, it is difficult to imagine that nuclear strategic models could realistically invoke

the same justification. This is because the stakes of failure are too high. In the context of deterrence, for instance, the "sorting mechanism" would presumably be nuclear annihilation, which is far too blunt and final an instrument to be effective in fashioning rationality.

Even if nuclear strategists could make a case for rationality on these grounds, however, then the point would still be moot insofar as we are interested in the applicability of cognitive psychology. This is because cognitive psychology, again, has little application to models that assume rationality on this basis.

Cognitive psychology has no bearing on "selection"-based models of rational behavior because such models are operationally indifferent to the psychology of the actors involved. The truth of this becomes clear if we consider that such models do not even require that actors *have* a psychology.[21] Take, for example, evolutionary biology, which has had great success using game-theoretical (i.e. rational actor) models justified on these grounds. When biologists explain the behaviors of fish with reference to the (evolutionary) "rationality" of those behaviors,[22] they are not saying that fish make consciously rational decisions.[23] They are asserting, rather, that certain instincts proved more "rational" when aggregated over time — in the sense that they conferred slight but meaningful advantages to reproductive success — and, as a consequence, were favored by natural selection.

iv. Rationality as Emergent Mean or Mode

A final justification for invoking rational actors in theoretical models is to focus on trends rather than individuals, and to assume that actors behave rationally "on average" even if

they can be capricious as individuals. This claim can more accurately be thought of as two distinct justifications with different mechanisms, which are grouped together here because they are functionally equivalent in their relationships to nuclear strategic models and cognitive psychology. These are: (a) to suppose that individual behaviors cluster around a rational *mode*, and (b) to suppose that individual behaviors converge on a rational *mean*.

To say that behaviors cluster around a rational *mode* is to assume that more people make the rational choice than any other. Consider, for example, a group of individuals shopping for a new car and reduced to choosing between two options that are identical in all meaningful ways except their price. In these circumstances some individuals might choose the more expensive — economically "irrational" — option for personal and circumstantial reasons (perhaps it is being sold by a friend or relative). Yet it is nevertheless reasonable for economists to expect the majority of shoppers to choose the cheaper — economically "rational" — option, which will therefore become the modal choice.

To say that behaviors converge on a rational *mean* is to assume that the irrational behaviors of individuals cancel each other out in the aggregate, making the decisions of the collective (expressed, for instance, in the price of a stock) more rational than those of the individuals of which it is composed. Economists routinely make this assumption. A consequential example would be Hayek's argument that free markets efficiently integrate individual behaviors in ways that reveal optimal (and thus "rational") distributions and prices.[24]

It is primarily in this context — where models are basing a rational actor assumption on the justification that rationality lies in means and modes — that economists have found cognitive psychology to be valuable. As outlined at the top of this chapter, the core insight of behavioral economics is that biases often *do not* cancel each other out or cluster around a rational mode, but can instead converge towards "predictably irrational" outcomes.[25] Identifying such irrationalities, along with their intensity and the circumstances in which they occur, has allowed economists to hone their models and interventions in many ways — informing the designs of everything from retirement schemes to gambling machines.[26]

Crucially, however, it is difficult to imagine that the models of nuclear strategists can be justified (and therefore modified) on these grounds. The key point here is that these justifications for assuming rationality only apply when theorists are interested in means and modes (in the form of trends, for example, or stock prices), and are relatively indifferent to isolated datapoints (such as individual choices or bids), which they assume to be messy and unpredictable. Yet individual incidents *matter* in nuclear strategic models, which cannot afford to treat them as inherently unpredictable. In this sphere the fate of civilization can hinge on a single decision by a single actor: be it a parliament, a president, or even a mid-level functionary like Vasili Arkhipov[27] or Stanislav Petrov.[28] And where an isolated datapoint could mean a war that rendered all models and stratagems moot forever, it seems illogical for theorists to dismiss specific interactions as "noise" and look past them to a "signal" in their aggregate.[29]

Rationality Revisited

This is all to say that cognitive psychology is only valuable to models that make specific assumptions about rationality, and that the models of nuclear strategists (unlike those of many economists) cannot afford make these assumptions (See table 9.1). It follows from this that strategic models should have little use for cognitive psychology. Indeed, nuclear strategists might gain more from doubling down on their rationality assumptions than from questioning them.

-- Table 9.1 about here --

Although we believe the essential logic of this argument to be sound, we freely concede that it deserves more nuance than can be afforded here. Let us therefore close this section with two caveats. The first is to note that some of the models deemed irrelevant to cognitive psychology above, might, at the margins, find some value in explorations of human irrationality. Where models are intended to be normatively prescriptive, for example, there might be scope for cognitive psychology to highlight areas were people are most in need of guidance. The second is to note that, even if strategic models are unlikely to benefit from insights into cognitive biases, strategic thinkers might still gain from looking to psychology more broadly. Prominent cold warriors have come to believe be that rational actor models are poor tools for guiding nuclear strategy in a world of "rogue states" and terrorist organizations.[30] In such circumstances it may be valuable to abandon game-theoretic models entirely and to look for strategic insights in aspects of psychology that economists have largely ignored.

Part 2: On Actors

Methodological Individualism and the Problem of Aggregates

In the previous section we examined the different justifications that rational choice theorists invoke for ascribing rationality to economic and strategic actors. In this section we will more directly examine the actors to which they ascribe that rationality.

In nuclear strategic discourse, theorists apply theories of rational choice to many different types of actors, about which they require little substantive knowledge.[31] As long as the actor can be treated as unitary, rationality is equally as applicable to organizations, such as nation-states and firms, as it is to individuals. Behavioral economics, by contrast, is rooted in experimental fields that exclusively take individual human cognition to be their object of inquiry.[32] As such, its findings are not only specific to the behavior of individual beings, but often also linked to the underlying biology of the human brain. In *Thinking, Fast and Slow*, for instance, Kahneman argues that people often depend on simple heuristics because slower and more deliberate thought processes require physiological effort and resource expenditure.[33] (So it is, for example, that scientists have found that people are more likely to engage in rigorous analysis instead of making a snap judgment after rather than before lunch.) The fact that this approach is based on the study of individual human behavior, however, raises question about what applicability, if any, these findings have to the collective and aggregate actors often invoked by nuclear strategists.

That there is a problem with applying findings based in methodological individualism to collectives and aggregates is widely recognized.[34] In the conclusion to a recent special issue on "The Behavioral Revolution and International Relations," for instance, Janice Stein argues that psychological explanations of international behavior are undermined by processes of aggregation. She concludes that the methodological

individualism of cognitive psychology creates an inference problem because "…theoretical propositions drawn from individual-level analysis do not move easily to 'higher-level' units such as states." [35] In a similar vein, economists Bruno Frey and Jana Gallus have proposed that aggregation processes may produce different outcomes than those observed at the individual level, and that the problem of understanding these differences is substantial enough to constitute a new field of research.[36]

Behavioral researchers have generated a number of specific insights that pertain to strategic planning and decision making, but have struggled to integrate these insights into a single, comprehensive framework.[37] This could be, at least in part, due to the fact that theories of deterrence and nonproliferation typically take the state as the primary actor, and treating a state as if it behaves like an individual is to commit what is known as a category mistake.

A category mistake, in its most general sense, is when someone mistakenly ascribes qualities to an object or entity that it clearly does not possess. The mid-twentieth century philosopher Gilbert Ryle introduced the term to describe the false equation of two entities that do not share the same ontological status, or the misattribution of characteristics to an entity that cannot, by definition, exhibit such characteristics. He offers several examples, such as the following, which revolves around the ontological status of the University:

> A foreigner visiting Oxford or Cambridge for the first time is shown a number of colleges, libraries, playing fields, museums, scientific departments and administrative offices. He then asks 'But where is the University? I have seen where the members of the Colleges live, where the Registrar works, where the

> scientists experiment and the rest. But I have not yet seen the University in which reside and work the members of your University.'[38]

The category mistake here is to expect that the University will be a discrete building rather than a collective entity that is the association of these discrete parts. States are made up of individuals, but to treat them as if they possessed the characteristics of a human being is to make the same mistake as Ryle's tourist. The state does not exhibit the characteristics of a human body any more than the University exists as a single building.

Avoiding category mistakes is one of the challenges of applying the insights of behavioral economics to questions of nuclear deterrence and nonproliferation. Unlike the "rational actor" assumption, which, in some circumstances, is equally applicable to states, organizations, and individuals, the findings of behavioral economics are based on observations of human behavior and rooted in the physical processes of the brain. The experimental basis of those observations, meanwhile, grants them a different ontological status than that of an *a priori* assumption. But whereas an assumption is easily transported from one domain to another, the substantive findings of psychological and neurobiological experiments are not so readily transferred. Claiming that a state might be predictably irrational in the same way as a person is a significant and problematic leap. So it is that the "problem of aggregate actors" offers a second reason to doubt there is a straightforward application of findings from behavioral economics to the realm of nuclear strategy.

Towards a Reconciliation?

The complexities of collective behavior undoubtedly raise important questions for deterrence strategists looking to invoke behavioral insights. As with the discussion of rationality above, however, and probably to a greater degree, it is possible to envisage ways that nuclear strategists might grapple productively with the problems of applying cognitive psychology to collective actors. To this end, it is worth looking to other research programs that have faced the challenges of reconciling methodological individualism with aggregate subjects. There are at least two such examples: (i) the study of affect in International Relations; and (ii) the study of decision making in bureaucratic organizations. We will briefly examine each in turn.

i. Affect and the State

Scholars interested in the role of emotions in international relations contend with a problem similar to the one identified above.[39] Like cognitive biases, emotions are individual-level phenomena that are understood in relation to their biological basis, and which have a complex relationship to institutional actors. Although emotions such as anger, sympathy, and guilt are easily observable in the rhetoric of international interactions, the focus on state-level behavior in dominant theories of international relations excludes such emotions from consideration. For while it is easy to observe emotion at the individual level in the behavior of leaders and members of the public, it is difficult to theorize it as a driver of state behavior.

Political scientists Todd Hall and Andrew Ross have tackled the problem of linking micro-level emotions to macro-level behaviors. They theorize three pathways for collective affective experience: bottom-up through shared concerns and dispositions;

horizontally across individuals through contagion; and top-down as a result of social harmonizing processes.[40] In exploring these pathways they describe "affective waves," which occur in response to an event that elicits strong emotions across individuals, the shared experience of which further intensifies the response and becomes capable of overriding pre-existing goals and concerns.[41] They also speak of "emergent collective solidarities," wherein the political identities of groups are shaped through shared emotional responses.[42] (For an illustration of these collective dynamics we might look to the Arab Spring, where a shared response to Mohamed Bouazizi's self-immolation in Tunisia inspired a series of protests across the Middle East, resulting in major social uprisings.)

By providing a connection between the individual experience of emotion and the collective phenomena of "affective waves" and "emergent collective solidarities," Hall and Ross show how theorists might grapple with the relationship between individual-level phenomena and state-level behavior. It is not implausible, moreover, that the pathways they outline might offer a tool for ascribing biases to collective actors. For while some of the biases and heuristics identified by behavioral economists are cognitively driven rather than rooted in affective experience,[43] others have a clear emotional basis. Take, for example, loss aversion (a stronger desire to avoid losses over realizing equivalent gains) and the endowment effect (a tendency to overvalue goods we already own), both of which are readily identifiable as being linked to feelings of attachment.[44] One could easily imagine pathways and collective-level phenomena through which loss aversion and the endowment effect make it easier for a state to forego

building nuclear weapons than for it to give them up having crossed the nuclear threshold.[45]

ii. Bureaucracies

An alternative approach that scholars have taken to the problem of extrapolating from individual to collective actors is to take an explicitly "organizational" approach, and ask how institutional structures interact with individual-level decision making biases and heuristics. This approach recognizes that key individuals (with all their biases) can have real agency even within large organizations, and seeks to explore how, and where, that agency is exercised.

The organizations that are most pertinent from the perspective of nuclear strategy are the state bureaucracies that frame and process key decisions (about whether or not to send information about a possible attack up the chain of command, for example.)[46] The question of how institutional incentives interact with biases and heuristics to shape decisions in bureaucratic contexts could potentially constitute an important avenue of behavioral research.[47]

Conclusion

Sigmund Freud is often said to have deemed the Irish to be entirely immune to psychology. It is unlikely that the same could be said of *homo atomicus,* yet the insights of psychologists undoubtedly have a complicated relationship to nuclear strategy. In this chapter we have presented two critiques of the "behavioral revolution" as it applies to nuclear strategic thought.

In the first we evaluated applying the findings of cognitive psychology in light of justifications that theorists invoke for assuming rational actors. Cognitive psychologists have shown that people have biases, which, in the aggregate, lead to consistent and predictable outcomes. Economists — who often build models on the assumption that people behave rationally in the aggregate and isolated events can profitably be ignored — have found this insight to be useful. Nuclear strategists, by contrast, do not build models on assumptions about the aggregate behavior of groups and the insignificance of isolated cases. They build models that speak to isolated cases, and do so on the belief that those models can inform and/or usefully explain the behavior of informed actors. As such, we argued, they are less likely to either want or need to correct for unconscious biases.

In the second we addressed the challenge of applying individual-level findings to state-level behaviors. Unlike the rational actor assumption, which moves relatively easily across different levels of analysis, we argued, simply transferring the results of cognitive psychological experiments to more complex social actors is not so straightforward. The substantive claims that cognitive psychologists make about heuristics and biases are rooted in the biological processes of the human brain. Organizations do not share this physiology, however, and we contended that to treat more complex social actors, such as states, as if they would display the same biases as an individual is to commit a category mistake.

Neither critique is definitive, as we explain, and it is possible to imagine research programs that attempt to navigate the limitations they describe. We do, however, believe that both are foundational and far reaching. There are compelling reasons to believe that

the insights of behavioral psychology will not map easily or straightforwardly onto the problems of nuclear strategy; it is important that scholars remain alive to this.

JUSTIFICATION FOR RATIONAL ACTORS	Relevant to nuclear strategic models?	Amenable to behavioral adjustment?
i. Normative prescription	Yes	No
ii. Special Circumstances	Yes	No
iii. Systematic Selection	No	No
iv. Emergent mean/mode	No	Yes

Table 9.1: Justifications for assuming rationality and their implications for a behavioral approach to nuclear strategy

Notes

[1] e.g. Thomas Schelling, *Arms and Influence (*New Haven: Yale University Press, 1966) .

[2] *Homo economicus* was never entirely uncontroversial, however, and predictions of his demise have been around for decades. See, e.g., Joseph Persky, "Retrospectives: The Ethology of Homo Economicus," *The Journal of Economic Perspectives* 9, no. 2 (Spring 1995).

[3] Testifying before Congress, former United Sates Federal Reserve Chairman, Alan Greenspan, has admitted to making foundational mistakes about rational actors in the financial system. See, e.g. Andrew Clark and Jill Treanor, "Greenspan - I was wrong about the economy. Sort of," *The Guardian*, Oct.28, 2008, https://www.theguardian.com/business/2008/oct/24/economics-creditcrunch-federal-reserve-greenspan (accessed 06/02/018).

[4] This is conveyed clearly in the title of Dan Ariely's book, *Predictably Irrational* (London: Harper Collins, 2008).

[5] Our intention is to speak specifically to cognitive psychology's relationship to nuclear strategy, not to its relationship with international relations or political science more broadly. Many of our points will apply across these contexts, but the unique nature of nuclear weapons also creates specificities that will be significant to our argument.

[6] Charles Brockden Brown, "Somnambulism, A Fragment," in *American Gothic: From Salem Witchcraft to H. P. Lovecraft*, ed. Charles L. Crow (Malden, MA: Wiley-Blackwell, 2013), p. 26.

[7] There are complex debates around the question of whether nuclear strategic models *should* be premised on rationality (especially given that its strategies routinely call for

displays of irrationality); see e.g. Patrick Morgan, *Deterrence Now* (Cambridge: Cambridge University Press, 2004), 42-79. Such questions are beyond the scope of this chapter, however, and in the following discussion we will limit ourselves to the question of whether (and how) rational agent theories of deterrence (such as they are) stand to gain from the insights about psychology that feature in behavioral economics.

[8] "Bounded rationality," a term coined by Herbert Simon, is the idea that individuals attempt to act rationally but are constrained by limited information, time, cognitive ability, etc. Herbert Simon, "A Behavioral Model of Rational Choice," in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting* (Wiley: New York 1957); Herbert Simon, "Human Nature in Politics: The Dialogue of Psychology with Political Science," *American Political Science Review* 79 (1985): 293–304.

[9] In discussing the behavioral revolution and its applicability to international relations, for example, Hafner-Burton et al. make an explicit point of excluding "normative" (as opposed to "positive") theories of human behavior. Emilie M. Hafner-Burton, Stephan Haggard, David A. Lake, and David G. Victor, "The Behavioral Revolution and International Relations," *International Organization* 71, issue S1 (Supplement 2017), S1–S31, p. 7.

[10] For further examples of rationality being invoked in this fashion by economists see, for instance, L. Savage, *The Foundations of Statistics*, 2nd ed. (New York: Dover Publications, [1956] 1972). This justification for rational actors is particularly prominant in fields like ethics and the philosophy of practical reason. See, for instance, R.

Cunningham, "Ethics and Game Theory: The Prisoner's Dilemma," *Papers on Non-Market Decision Making* 2 (1967): 11-26.

[11] Evidence for deterrence theory's normativity can be found in the fact that many of its core recommendations were highly counterintuitive, suggesting that they were only adopted because the theory itself was authoritative. See J. DeNardo, *The Amateur Strategist: Intuitive Deterrence Theories and the Politics of the Nuclear Arms Race* (Cambridge: Cambridge University Press, 1995).

[12] Morgan, *Deterrence Now*, 42.

[13] Although, in principle, it might be useful to understand the common biases that lead people astray--to ascertain where explicit rules would be most useful, for instance.

[14] It is worth noting here that this justification for assuming rational actors can easily be combined with the previous justification. This is simply to observe that in circumstances where the stakes of a decision are very high, it becomes more likely that actors will model optimal solutions, and more likely that they will adhere to the prescriptions of those models.

[15] Indeed, many institutions are actively designed to mitigate biases — for example by facilitating the expression of dissent. See Elizabeth Saunders, "No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making," *International Organization* 71, issue S! (Supplement 2017): S219–S247. See also: M. Taylor, "When Rationality Fails," in J. Friedman, (ed.), *The Rational Choice Controversy: Economic Models of Politics Reconsidered* (New Haven: Yale University Press, 1995): 223–234, J. Ferejohn & D. Satz, "Unification, Universalism, and Rational

Choice Theory," in *ibid*.; *S.* Kelley, "The Promises and Limitations of Rational Choice Theory," in *ibid*.

[16] Morgan, *Deterrence Now*, 42, 66; J. Mueller, *Retreat from Doomsday: The Obsolescence of Major War* (New York: Basic Books, 1989).

[17] In Scott Sagan & Kenneth Waltz, *The Spread of Nuclear Weapons: A Debate* (New York: W. W. Norton, 1995), p. 98.

[18] See for instance Morgan, *Deterrence Now*, 70-72.

[19] Such models invariably define "rationality" in relation to themselves. Such that it makes sense to speak of a "market rationality" which might differ significantly from any intuitive sense of rationality.

[20] e.g. L. Blume & D. Easley, "If You're so Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets," *Econometrica* 74 no. 4 (2006): 929–966.

[21] We might say that the agency in such models resides in the system rather than the actors themselves.

[22] See for instance Peter Buston & A. Zink, "Reproductive Skew and the Evolution of Conflict Resolution: A Synthesis of Transactional and Tug-of-War Models," *Behavioral Ecology* 20 (2009): 672-684.

[23] Nor are they saying that every "irrational" fish fails to prosper, only a disproportionate number in the aggregate (on this, see our next point).

[24] Friedrich Hayek, "The Use of Knowledge in Society," *American Economic Review* 35, no. 4 (1945): 519-30. This assumption also has some surprising applications beyond markets. Consider, for example, a group of individuals trying to guess the weight of an ox

at a country fair. Some guess too high and others too low, but if every guess is added together and then divided by the number of guesses to obtain a mean, then that mean tends to be closer to the correct weight than any of the individual guesses (to the extent that we might want to call it the most "rational" guess). This counterintuitive phenomenon has been well documented in a range of contexts, having been first recognized by Francis Galton in 1909 (in the context of guessing the weight of an ox). James Surowiecki, *The Wisdom of Crowds* (New York: Random House, 2004).

[25] Ariely, *Predictably Irrational*.

[26] Richard Thaler & Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness.* (New Haven: Yale University Press, 2008).

[27] See for instance Noam Chomsky, *Hegemony or Survival: America's Quest for Global Dominance* (New York: Owl Books, 2004) : 74

[28] See for instance Tony Long, "Sept. 26, 1983: The Man Who Saved the World by Doing … Nothing," *Wired*, September 26. 2007, https://www.wired.com/2007/09/dayintech-0926-2/ (accessed 18 April 2017).

[29] It should be recognized that in some respects nuclear strategy is undoubtedly an iterated game with emergent norms. Nye, for example, argues that the U.S.-Soviet relationship became more stable over time as parties learned from each other. But such learning cannot be the fundamental justification for invoking rational actors, as rational theories of strategy assume actors "get it right" from the start and do not need to undergo a learning process. J. Nye, "Nuclear Learning and US–Soviet Security Regimes," *International Organization* 41, no. 3 (1987): 371–402.

[30] See Philip Taubman, *The Partnership: Five Cold Warriors and their Quest to Ban the Bomb* (New York: Harper Perennial, 2013).

[31] A notable exception is the literature on strategic culture. See for instance Jeffrey S. Lantis, "Strategic Culture and Tailored Deterrence: Bridging the Gap between Theory and Practice," *Contemporary Security Policy* 30, no. 3 (2009).

[32] In *Thinking, Fast and Slow* (New York: Farrar, Strauss and Giroux, 2011), Daniel Kahneman references both the cognitive psychology and cognitive neuroscience literature extensively.

[33] The nervous system consumes more glucose than other systems in the body and cognitive function is especially sensitive to its availability. See Matthew T. Gailliot, et al., "Self-Control Relies on Glucose as a Limited Energy Source: Willpower is More than a Metaphor," *Journal of Personality and Social Psychology* 92, no. 2 (2007): 325-336. See also Kahneman, *Thinking, Fast and Slow*, 43.

[34] See for instance Robert Powell, "Research Bets and Behavioral IR," *International Organization* 71, l issue S1 (Supplement 2017): S265-S277; Bruno S. Frey and Jana Gallus, "Aggregate Effects of Behavioral Anomalies: A New Research Area," *Economics Discussion Papers*, No. 2013-51 (Kiel Institute for the World Economy, 2013), http://www.economics-ejournal.org/ economics/discussionpapers/2013-51. For a related discussion of a research program for the application of findings about emotions to aggregates see Todd Hall and Andrew Ross, "Affective Politics after 9/11," *International Organization* 69 (Fall 2015), pp. 847–879; Todd Hall, *Emotional Diplomacy: Official Emotion on the International Stage* (Ithaca: Cornell University Press, 2015).

[35] Janice Stein, "The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics," *International Organization* 71, issue S1 (Supplement 2017), pp. S249–S263, quote at S255.

[36] Frey and Gallus, "Aggregate Effects of Behavioral Anomalies."

[37] This was also the finding of a recent special issue in *International Organization*. See Hafner-Burton, et al., "The Behavioral Revolution and International Relations," 1.

[38] Gilbert Ryle, *The Concept of Mind*, 60th Anniversary Edition (London and New York: Routledge, 2009), p. 6.

[39] On emotion and international relations see: Hall and Ross, "Affective Politics after 9/11"; Hall, *Emotional Diplomacy: Official Emotion on the International Stage*; Jonathan Mercer, "Human Nature and the First Image: Emotion in International Politics," *Journal of International Relations and Development* 9, no. 3 (2006): 288–303; Jonathan Mercer, "Emotional Beliefs," *International Organization* 64, no. 1 (2010):1–31; Jonathan Mercer, "Emotion and Strategy in the Korean War," *International Organization* 67, no. 2 (2013): 221–52.

[40] Hall and Ross, "Affective Politics after 9/11," 848.

[41] Ibid., 859.

[42] Ibid.

[43]As an example, anchoring effects (a tendency to over-rely on early information) are primarily cognitive in nature because the radical contingency of the anchor precludes their effects from being motivated by emotional experience.

[44] A common example of a state-level endowment effect that likely operated through these emotive pathways is the Falklands War in which the UK fought a costly war to

preserve possession of a resource-poor remote South Atlantic archipelago. By arousing nationalistic sentiments Prime Minister Margaret Thatcher was able to generate the support for defending the islands against Argentinian attack, though making a similar case to occupy the islands were they not already under British administration would be hard to imagine. See Frey and Gallus, "Aggregate Effects of Behavioral Anomalies," pp. 3-4.

[45] See also the chapter by Etel Solingen in this volume, which makes a similar argument for why sanctions and positive inducements are more likely to persuade a state to give up a nuclear weapons program the earlier the stage of the development program at which these tools are applied.

[46] See for instance Janice Gross Stein and Morielle I. Lotan, "Disabling Deterrence and Preventing War: Decision Making at the End of the Nuclear Chain" in this edited volume.

[47] By taking Kahneman's model of "two systems" within the brain, for example (i.e. a "fast" associative system of short-cuts, and a "slow," analytical system), we might imagine a form of analysis that maps these systems onto different bureaucratic structures. For instance, they might match up with different sections of the political and military bureaucracy that are responsible for rapid response and longer-term planning, respectively. (This approach is inspired by Slovoj Zizek's recovery of Lacanian psychoanalysis for the study of political phenomenon. See Slavoj Zizek, *The Sublime Object of Ideology* [New York and London: Verso, 1989]). In this context, one could imagine that systematic biases (manifest at the collective rather than the individual level) might be more prevalent in structures that work with shorter timeframes.