

H-Diplo | ISSF Roundtable 13-2

issforum.org

Dominic D.P. Johnson. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics*. Princeton: Princeton University Press, 2020. ISBN: 9780691137452 (hardcover, \$27.95).

22 October 2021 | <https://issforum.org/to/ir13-2>

Editor: Diane Labrosse | Commissioning Editor: Manjari Chatterjee Miller | Production Editor: George Fujii

Contents

Introduction by Valerie M. Hudson, Texas A&M University.....	2
Review by Joshua D. Kertzer, Harvard University.....	5
Review by Marika Landau-Wells, University of California, Berkeley.....	9
Review by Janice Stein, University of Toronto	12
Response by Dominic D. P. Johnson, University of Oxford	16

INTRODUCTION BY VALERIE M. HUDSON, TEXAS A&M UNIVERSITY

Evolutionary Biases in Foreign Policy Making

Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics by Dominic D.P. Johnson is a welcome addition to the literature on Foreign Policy Analysis (FPA). The study of cognitive biases has a long and rich history within FPA, with classics penned by luminaries such as Robert Jervis, Richards Heuer, Yaacov Vertzberger, Philip Tetlock, and Yuen Foong Khong, alongside pioneering work among behavioral economists such as Daniel Kahneman and Richard Thaler.¹ The ‘moral’ of much of this work has been that cognitive biases are liabilities that trip up rational decision-making, and foreign policy decisionmakers must work diligently to mitigate their influence.

How refreshing and provocative, then, is Johnson’s thesis that cognitive biases have been shaped by evolution to provide us with an *advantage* in decision-making when the stakes are high, such as in matters of war and conflict. Our first-blush reaction in such a situation, based on these biases, may actually be a good guide to action, especially when the circumstances are fraught with uncertainty. Rather than thinking of them as biases, we might better see them as evolutionarily-tested heuristics that speed decision-making and steer it in the direction of a greater chance of survival, for “evolution has had millions of years of trial and error to craft effective methods of making decisions under uncertainty with limited time and information” (19). Johnson proposes that “biases may make us better at setting ambitious goals, building coalitions and alliances, bargaining effectively, sending credible signals, maintaining resolve, and persevering in the face of daunting challenges, and they may make us more formidable when it comes to conflict, deterrence, coercion, crisis, brinkmanship, and war” (6).

The three biases Johnson explores are overconfidence, the fundamental attribution ‘error’ (i.e., imputing purposefulness to the action of others), and in-group bias. While one’s perceptions colored by these biases may at times be objectively wrong, evolution suggests that over time they will be more right than wrong, for these biases help us avoid making the most costly errors. As Johnson notes, while sometimes we mistake a stick for a snake, we tend not to mistake a snake for a stick, a cognitive bias that is evolutionarily useful. Furthermore, even if at times a bias may be wrong in some sense, it may still be useful, such as serving to solidify domestic support for war through in-group bias. While Johnson does not discount the idea that sometimes these cognitive biases will lead to disaster, he suggests that over the long term, the biases will bolster our chances of survival—which, after all, is evolution’s sole yardstick. Whether they make the world a more conflictual place as well is beside evolution’s point. Johnson’s recommendation is that “instead of finding ways to avoid or suppress cognitive biases, we should look for ways to channel them so they can better work their magic” (15). To use a Star Trek image, Captain Kirk, not Commander Spock, then, is what is needed.

Johnson’s book offers a needed antidote to the current trend of ideological opposition to any mention of evolution in human affairs, wherein social science “training has burned into [scholars] the idea that evolutionary explanations of human behavior are fundamentally flawed, hopelessly deterministic, or morally reproachable. Such views may come to represent a lost generation of social science—scholarship left behind because it fought science as an enemy rather than engaging with it as an interdisciplinary endeavor and pursuing the consilience of knowledge” (37). A social science that cannot enter into a dialogue with the life sciences is almost by definition an intellectual dead-end. Johnson rightly asserts, “While we may not

¹ Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976); Richards Heuer, Richards, *The Psychology of Intelligence Analysis* (Washington, D.C.: Government Printing Office, 1999); Yaacov Vertzberger, *The World in Their Minds: Information Processing, Cognition, and Perception in Foreign Policy Decisionmaking* (Stanford: Stanford University Press, 1990); Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton: Princeton University Press, 2006); Yuen Foong Khong, 1992. *Analogies at War: Korean, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton: Princeton University Press, 1992); Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982); Richard Thaler, “Mental Accounting Matters.” in Kahneman and Tversky, eds., *Choices, Values, and Frames* (Cambridge: Cambridge University Press, 2000).

like everything we find out about human nature, we should like to understand it. We have a far better chance of solving perennial problems of security, cooperation, and justice by comprehending human nature than by ignoring or denying it” (45). Amen to that.

Johnson continues his investigation of these biases by means of illustrative historical cases demonstrating how they did play into (or would have played into) a successful outcome. Some of these cases are more convincing than others (the FAE case seems fairly tautological). I was struck by two unaccountable lacunae in the analysis. The first is that all three biases are deeply gendered: women are not, generally speaking, overconfident and are, generally speaking, far less inclined to intensive “chest-pounding” (183) nationalism; women are also arguably more vigilant about others’ motives, especially those of men. The importance of these particular biases in the context of war and conflict may be due to the overwhelming preponderance of male leaders, and the volume does not inquire into that proposition, nor does it ask what the asserted linkage of these biases with success means for women and women’s leadership on the national stage.

The second lacuna is the lack of reference to the subfield of Foreign Policy Analysis (FPA). While occasionally Johnson alludes to the work of “neoclassical realists and constructivists” (138), one would never know that there was a vibrant subfield of FPA from reading this book whose subject is a core level of analysis for FPA scholars.² At one point, Johnson notes, “It is not clear where the intellectual progression [of the field of international politics] would lead next, but it seems likely there will be a rebalancing such that we rediscover the role of individuals and come to appreciate the important of both dispositional and situational factors” (139). In fact, FPA has been leading out with theoretical and empirical work in this vein since the 1950s, so this is an unaccountable omission.

Three able reviewers have contributed to this roundtable: Joshua Kertzer, Marika Landau-Wells, and Janice Stein. All are appreciative of the work, seeing in Johnson’s thesis an important contribution to the literature, both in terms of bringing in evolutionary insights as well as pushing back against the idea that biases are always disadvantageous. The reviewers each point out some areas needing greater attention. Stein notes that while Johnson’s focus is on the level of the individual leader, evolution is not concerned with individual success, but with species success defined in terms of reproduction. What reproduction means, then, in terms of the nation-state is not a discussion with which Johnson engages, which renders the evolutionary analysis problematic (especially in a day of weapons of mass destruction). In his response, Johnson posits that reproduction need not enter into an explanation based on the evolutionary shaping of the human mind, but then that muddies the concept of success and how it is to be gauged. Stein also asks a foundational question that is not answered by Johnson: Is it in fact the case that biases produce greater success on average over time? And given that the most important phenomena in international affairs, such as wars, are relatively rare, what does success *on average* mean?

Landau-Wells adds to the discussion by noting that the relatively tight conceptualizations of the three cognitive biases undergo considerable stretching in the case studies. For example, she notes that the fundamental attribution error (FAE) broadens in Johnson’s case study of British Prime Minister Neville Chamberlain to become the larger concept of threat perception. However, the FAE refers to occasions when we are *mistaken* about the other’s motives, seeing their behavior as purposive rather than determined by the circumstances when in fact they are determined by circumstances. She rightly asks why making such a mistake would lead to better detection of genuine threats. Landau-Wells also notes the problem inherent in Johnson’s research: how can we contrast biased judgment against accurate judgment if even historians cannot settle on an ‘accurate’ picture of historical circumstances? Echoing Stein, how then do we really know a bias was in fact involved?

Kertzer has several interesting observations to add as well. Noting the uniqueness of Prime Minister Neville Chamberlain and President George Washington, and how many of their contemporaries would have made different decisions if they had been the leader at the time (for example, Prime Minister Winston Churchill), does raise the theoretical issue of how much

² For an overview of the literature, see Valerie M. Hudson and Benjamin S. Day, *Foreign Policy Analysis: Classic and Contemporary Theory*, 3rd ed. (New York: Rowman and Littlefield, 2019); see also the journal *Foreign Policy Analysis*.

causality is to be imputed to variation at the level of the individual leader. Kertzer argues that if such variation is the causal crux, and not the biases, then Johnson has but moved the theoretical goalposts. Kertzer also notes that Johnson pays insufficient attention to the competing conflict arena of domestic politics, and how those domestic contests might vie with international affairs as the top threat in a leader's mind.

All three reviewers note that the question of how these cognitive biases are to be moderated, or "channeled so they can better work their magic" (15) is quite unclear even after a thorough reading of Johnson's book.

In sum, then, Johnson has made an important theoretical move at the intersection of evolution, cognition, and international affairs. While his book may not be the final say on that point of intersection, it provides a foundation for others to build upon. I plan on recommending it to my graduate students this year.

Participants:

Dominic D.P. Johnson received a DPhil from Oxford University in evolutionary biology, and a PhD from Geneva University in political science. Drawing on both disciplines, he is interested in how new research on evolution, biology and human nature is challenging theories of international relations, conflict, and cooperation. His previous books include *Failing to Win: Perceptions of Victory and Defeat in International Politics* (Harvard University Press, 2006), with Dominic Tierney, which examines how and why popular misperceptions commonly create undeserved victories or defeats in international wars and crises, and *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Harvard University Press, 2004), which argues that common psychological biases to maintain overly positive images of our capabilities, our control over events, and the future, play a key role in the causes of war. His current work focuses on the role of evolutionary dynamics, evolutionary psychology, and religion in human conflict and cooperation.

Valerie M. Hudson is a University Distinguished Professor in the Department of International Affairs at The Bush School of government and Public Service at Texas A&M University, where she directs the Program on Women, Peace, and Security. Hudson has been named a Distinguished Scholar of Foreign Policy Analysis, and has authored a popular textbook of the subfield, entitled *Foreign Policy Analysis: Classic and Contemporary Theory*.

Joshua D. Kertzer is Professor of Government at Harvard University. His research specializes in the intersection of international security, political psychology, foreign policy, and public opinion. He is the author of *Resolve in International Politics* (Princeton University Press, 2016), along with articles appearing in a variety of academic journals, including the *American Journal of Political Science*, *Annual Review of Political Science*, *British Journal of Political Science*, *Conflict Management and Peace Science*, *International Organization*, *International Studies Quarterly*, *Journal of Conflict Resolution*, *Journal of Politics*, and *World Politics*.

Marika Landau-Wells is an Assistant Professor in the Travers Department of Political Science at the University of California, Berkeley. She received her Ph.D. in political Science from the Massachusetts Institute of Technology where she was also a Postdoctoral Research Fellow in the Department of Brain and Cognitive Sciences.

Janice Gross Stein is University Professor and Belzberg Professor of Conflict Management and Founding Director of the Munk School of Global Affairs and Public Policy at the University of Toronto. Her most recent articles are "The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics," *International Organization* 71 (2017): 249-263, and Janice Gross Stein and Ron Levi, "Testing Deterrence by Denial: Experimental Results from Criminology," *Studies in Conflict and Terrorism*, forthcoming 2021.

 REVIEW BY JOSHUA D. KERTZER, HARVARD UNIVERSITY

Dominic Johnson's *Strategic Instincts* is a remarkable book that seeks to overturn much of the conventional wisdom about cognitive biases in international politics. Against critics who view heuristics as deviations from rationality, and who unfairly reduce political psychology to the study of how decision-makers make mistakes,¹ Johnson builds on a growing body of scholarship in evolutionary psychology to argue that cognitive biases do more good than harm.² These tendencies that seem puzzling from the perspective of rational choice theory are actually adaptations crafted by natural selection, and which continue to serve adaptive functions in international politics today. In this sense, *Strategic Instincts* does more than merely preach the gospel of Leda Cosmides, John Tooby, and Gerd Gigerenzer,³ or push back against political scientists who believe in evolution but like to think it only works from the neck down. Crucially, the book shows how the laundry list of biases that are frequently invoked to explain failures in foreign policy decision-making – focusing here on overconfidence, the fundamental attribution error, and in-group/out-group biases – can also explain successes in IR as well. These biases are not unvarnished goods, to be sure: excessive overconfidence can lead to hubris, unbridled suspicion to paranoia, and inordinate intergroup bias to prejudice and discrimination. But in a world characterized by incomplete information, and where the costs of false negatives exceed the cost of false positives, it is better on average for our cognitive architecture to predispose us for false alarms than alarms that fail to go off at all.

The central contention of psychologists' claims that the behavioral sciences have a "bias bias"⁴ is that in order to gauge if a bias is functional you need to test it in the wild rather than in the lab. One of the real strengths of the book, then, is the rich set of case studies and dazzling array of illustrative examples it presents, which cover an enormous amount of ground, from the American revolution to the Munich Crisis, to the war in the Pacific. As Robert Jervis, Jack Levy, and others have noted,⁵ studying biases or misperceptions in the wild is challenging because of the absence of a neutral reference point: when is someone overconfident, versus appropriately so? When are decision-makers overattributing dispositional causes to others' behavior, and when are they accurately doing so? When do we know whether the level of intergroup bias is merely "moderate"? Johnson's empirical strategy here – both tracing perceptions over time, and drawing comparisons across observers – makes a noteworthy contribution in its own right, and the book a provocative and important read.

Strategic Instincts pushes us to consider a number of questions, including about evolutionary approaches themselves. Evolutionary theories are often seen as having two major appeals, both of which are invoked in this book. First, they are

¹ Joshua D. Kertzer and Dustin Tingley, "Political Psychology in International Relations," *Annual Review of Political Science* 21 (2018): 332, DOI: <https://doi.org/10.1146/annurev-polisci-041916-020042>, Kertzer, Jonathan Renshon, and Keren Yarhi-Milo, "How Observers Assess Resolve", *British Journal of Political Science* 51:1 (2021): 308-330, DOI: <https://doi.org/10.1017/S0007123418000595>

² Leda Cosmides and John Tooby, "Better than Rational: Evolutionary Psychology and the Invisible Hand," *American Economic Review* 84:2 (1994): 337-342, Gerd Gigerenzer and Wolfgang Gaissmaier, "Heuristic Decision Making," *Annual Review of Psychology* 62 (2011): 451-482, DOI: <https://doi.org/10.1146/annurev-psych-120709-145346>, Anthony C. Lopez, Rose McDermott and Michael Bang Petersen, "States in Mind: Evolution, Coalitional Psychology, and International Politics," *International Security* 36:2 (2011): 48-83, DOI: https://doi.org/10.1162/ISEC_a_00056

³ See, for example, Cosmides and John Tooby, "Better than Rational," Gigerenzer and Gaissmaier, "Heuristic Decision Making."

⁴ Henry Brighton and Gerd Gigerenzer, "The Bias Bias," *Journal of Business Research* 68:8 (2015): 1772-1784, DOI: <https://doi.org/10.1016/j.jbusres.2015.01.061>

⁵ Robert Jervis, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976) and Jack S. Levy, "Prospect Theory, Rational Choice, and International Relations," *International Studies Quarterly* 41:1 (1997): 87-112, DOI: <https://doi.org/10.1111/0020-8833.00034>

explanatory rather than merely descriptive, thereby turning psychology into more than just a laundry list of biases (32). Second, they are able to explain human universals – why these biases persist across cultures and contexts (181).

Yet one of the striking takeaways from the book's case studies is how much of international politics comes down to individual differences and individual-level heterogeneity rather than universality: it is not overconfidence in general that won the American revolution, but commander in chief George Washington's unusual confidence (making him the "indispensable man" – 101). British Prime Minister Neville Chamberlain, the book argues, would have correctly appreciated the threat posed by Nazi leader Adolf Hitler if he had been more susceptible to the fundamental attribution error (162-165) – a predisposition to which then-Conservative backbencher Winston Churchill was more highly prone. To their credit, evolutionary psychologists have wrestled with these questions, seeking to push a field that focuses more on species-typical adaptations to also account for individual differences,⁶ but there remains something striking about how much of the causal force in the book's accounts comes from leader-level variation – which is easier for us to describe than explain.⁷

The point is worth emphasizing for other reasons as well. The book provocatively notes that we wouldn't want to replace human leaders with robots because they lack our adaptive biases and thus would perform poorly (15). Importantly, however, leaders also perform differently from robots because they *vary* from another – and as these case studies adeptly show, leaders matter in international politics precisely because they aren't all the same. They are, as Johnson points out, "quirky" (13), displaying a heterogeneity that contradicts assumptions of efficient selection processes and overpowering socialization effects.

One of the interesting features of the book, then, is that although it contains a number of spirited critiques of neorealism, in other ways its evolutionary argument is in many ways sympathetic to it, both metatheoretically (evolutionary theories emphasizing human universals are helpful in understanding the nature of the international system, but not, as Kenneth Waltz put it, "why state X made a certain move last Tuesday") and substantively.⁸ The notion that international politics is a realm characterized by irreducible uncertainty, populated by functionally undifferentiated units, where it is more prudent for actors to assume the worst about one another's intentions, and for group members to unite in the face of external threat, is a world that many realists would find familiar.

The biases themselves also raise a number of questions. One of the interesting puzzles the book wrestles with is why overconfidence seems to have helped the Americans, but not the British, in the American revolutionary war. Fascinatingly, the book suggests it is partially due to an asymmetry in how overconfidence manifested itself on each side – the Americans were overconfident about their own side's strength, while the British were overconfident about their opponent's weakness (112) – but from a rationalist perspective this distinction is puzzling, raising questions about how we think about the role of

⁶ David M. Buss, "How Can Evolutionary Psychology Successfully Explain Personality and Individual Differences?" *Perspectives on Psychological Science* 4:4 (2009): 359-366, DOI: <https://doi.org/10.1111%2Fj.1745-6924.2009.01138.x>

⁷ Yarhi-Milo, *Who Fights for Reputation in International Politics? Leaders, Resolve, and the Use of Force* (Princeton: Princeton University Press, 2018), Michael C. Horowitz, Allan C. Stam and Cali M. Ellis, *Why Leaders Fight* (New York: Cambridge University Press, 2015), Peter K. Hatemi and Rose McDermott, "A Neurobiological Approach to Foreign Policy Analysis: Identifying Individual Differences in Political Violence," *Foreign Policy Analysis* 8:2 (2012): 111-129, DOI: <https://doi.org/10.1111/j.1743-8594.2011.00150.x>, and Brian C. Rathbun, *Reasoning of State: Realists, Romantics, and Rationality in International Relations* (Cambridge: Cambridge University Press, 2019).

⁸ Kenneth N. Waltz, *Theory of International Politics* (Boston: McGraw-Hill, 1979): 121.

mutual optimism in war, since in bargaining frameworks these two quantities should be complements of one another.⁹ Similarly, the book notes that while the American military commanders were overconfident, overconfidence on the British side was present in the aristocracy but not the field commanders. What explains this variation in overconfidence between different actors within a given side? If it is a function of differential information environments, what implications does this have for how we know an actor is overconfident, rather than appropriately confident given the information they have received?

The book's discussion of the fundamental attribution error (FAE) is also worth noting. In Lee Ross's classic formulation in social psychology, the FAE refers to the tendency "to underestimate the impact of situational factors and to overestimate the role of dispositional factors in controlling behavior."¹⁰ Following Jervis and Jonathan Mercer, who combined the FAE with self-serving and egocentric biases, political scientists often operationalize the FAE in terms of the tendency to view negative behavior as dispositionally motivated in particular,¹¹ and Johnson adopts a similar analytic move in the book, operationalizing the FAE in terms of actors assuming the worst about others' intentions (8); Chamberlain's reluctance to make dispositional attributions meant he underestimated Hitler's threat. Two points are relevant here. First, there is an interesting distinction between assuming others' behavior is dispositionally motivated, and overestimating the level of threat they pose. As Elizabeth Saunders notes in her work on leader beliefs in military interventions,¹² "internally-focused" leaders (who perceive threats to originate from the leadership of a country itself) don't necessarily perceive greater threats than "externally-focused" leaders do – just threats of a different type, which need to be addressed in a different type of way. Internally focused leaders, she argues, may in fact be *less* likely to intervene than externally focused ones.

Second, if the FAE is an adaptive trait intended to cause us to assume the worst about others' intentions in the face of threat, it can interact with other traits in theoretically interesting ways. Similar to Daniel Kahneman and Jonathan Renshon's arguments about why hawks win, Johnson notes that the three biases presented here should tend to push leaders towards greater hawkishness (276).¹³ Yet one can think of other biases that evolutionary scholars study that would have countervailing effects on the FAE. Hugo Mercier and Dan Sperber, for example, argue that motivated reasoning is an adaptive trait.¹⁴ If humans are predisposed to have heightened sensitivity to threats, but are also prone to seeing what they want to see, the two biases can point in opposing directions. This is the argument conservative critics make of America's

⁹ Alexandre Debs, "Mutual Optimism and War, and the Strategic Tensions of the July Crisis," *American Journal of Political Science* Forthcoming, DOI: <https://doi.org/10.1111/aips.12569>, and David Lindsey, "Mutual Optimism and Costly Conflict: The Case of Naval Battles in the Age of Sail," *Journal of Politics* 81:4 (2019): 1181-1196, DOI: <https://doi.org/10.1086/704221>

¹⁰ Lee Ross, "The intuitive psychologist and his shortcomings: Distortions in the attribution process," *Advances in Experimental Social Psychology* 10 (1977): 183, DOI: [https://doi.org/10.1016/S0065-2601\(08\)60357-3](https://doi.org/10.1016/S0065-2601(08)60357-3). See also Fritz Heider, *The Psychology of Interpersonal Relations* (New York: Wiley, 1958), Edward E. Jones and Keith E. Davis, "From Acts to Dispositions: The Attribution Process in Person Perception," *Advances in Experimental Social Psychology* 2 (1965): 219-266, DOI: [https://doi.org/10.1016/S0065-2601\(08\)60107-0](https://doi.org/10.1016/S0065-2601(08)60107-0), Harold H. Kelley, "Attribution theory in social psychology," "Attribution theory in social psychology," *Nebraska Symposium on Motivation* (Lincoln: University of Nebraska Press, 1967), and Daniel T. Gilbert and Patrick S. Malone, "The Correspondence Bias," *Psychological Bulletin* 117:1 (1995): 21-38, DOI: <https://psycnet.apa.org/doi/10.1037/0033-2909.117.1.21>

¹¹ Jervis, *Perception and Misperception in International Politics* and Jonathan Mercer, *Reputation and International Politics* (Princeton: Princeton University Press, 1996).

¹² Elizabeth N. Saunders, *Leaders at War: How Presidents Shape Military Interventions* (Ithaca: Cornell University Press, 2011).

¹³ Daniel Kahneman and Jonathan Renshon, "Why Hawks Win," *Foreign Policy* (January/February 2007): 34-38.

¹⁴ Hugo Mercier and Dan Sperber, "Why Do Humans Reason? Arguments for an Argumentative Theory," *Behavioral and Brain Sciences* 34:2 (2011): 57-111, DOI: <https://doi.org/10.1017/S0140525X10000968>.

engagement strategy with China: that American policymakers chose to overlook and play down Chinese provocations because the costs would otherwise be too severe.

Finally, I would be failing to enact my (thankfully, adaptive) ingroup favoritism if I didn't note that I sometimes wondered whether the book was too hard on political psychology – for example, characterizing it as “occupying the margins” of the discipline with its “powerful but often disregarded ideas” (31). Political psychologists as a matter of (strategic?) instinct have long written as if we are on the outside of the discipline looking in, but thanks to the pioneering work of scholars like Robert Jervis, Rose McDermott, Janice Gross Stein, Rick Herrmann, Jack Levy, and many others – including Johnson himself! – my sense is that the discipline has evolved. Perusing the top journals and book presses in the field reveals an abundance of psychological or behaviorally-informed work situated at the discipline's very core. Johnson's compelling book suggests that political psychologists' reflexive under-confidence may be worth shedding.

 REVIEW BY MARIKA LANDAU-WELLS, UNIVERSITY OF CALIFORNIA, BERKELEY

Scholars of international politics, and of conflict in particular, are often intrigued by rare events. Wars and crises stand out in relief against the day-to-day interactions between states, which makes them significant. But, when we look for the causes of rare events generally, we humans are biased, preferring to attribute rare events to rare causes.¹ Dominic Johnson's ambitious and insightful book, *Strategic Instincts*, offers an alternative perspective. *Strategic Instincts* demonstrates that ordinary psychological processes can provide compelling explanations for extraordinary events. The psychological processes which *Strategic Instincts* explores in depth are ordinary in two ways. First, they influence our behavior in daily life, not merely in times of crisis. Second, they are ordinary in that they operate in most humans, lay publics and world leaders, alike. Across four theoretical chapters and three case studies, Johnson makes the case that major events in international politics can result as much from the mundane, species-typical operations of the human mind as they can from exceptional pathologies or egregious errors.

Strategic Instincts also valuably pushes back against the tendency to study psychological mechanisms only when we want to explain *failures* in international politics. Instead, Johnson investigates the role that psychological biases play in *success*. He begins with a basic premise of evolutionary psychology: “[I]ntuitive judgment and decision-making mechanisms are the result of selection pressures we experienced in our evolutionary *past*, not in recent times” (38, emphasis in the original). In particular, *Strategic Instincts* centers on three biases in judgment and decision-making: the tendency towards overconfidence; the tendency to attribute behavior to dispositional factors rather than situational ones for our opponents, but not for ourselves (also known as the Fundamental Attribution Error, FAE); and the tendency to positively inflate characteristics of our ingroup(s) and to excessively derogate outgroups. Using a mixture of his own and others’ research, Johnson argues that these three tendencies should be regarded not as “biases” (an unambiguously negative term) but simply as “strategic instincts,” i.e., “rapid adaptive decision-making heuristics that we all have as human beings” (3). Importantly, Johnson notes that our biases “may in fact still be *adaptive* today, if they continue to bring benefits to their bearers—on average, over time” (40, emphasis in original). Thus, examining biases as contributors to success should be a fruitful endeavor. Given this perspective, the big questions at the heart of *Strategic Instincts* are: “whether and when [do] cognitive biases cause or promote success in the realm of international relations” (4)?

The answer that *Strategic Instincts* offers to the first of these questions – *whether* cognitive biases contribute to success – is clearly “yes.” To make this case, the book covers a significant amount of intellectual terrain. This breadth – in addition to the refreshing focus on psychology as a source of success – are two of the work’s major strengths. The first two chapters situate the book within the growing literature on the impact of human biology (broadly construed) on politics. The core logic of *Strategic Instincts* is rooted in evolutionary psychology, and, in particular on Error Management Theory.² The premise of EMT is that human decision-making strategies could be expected to have a bias – in the sense of favoring particular types of judgments regardless of the available information – if those judgments helped avoid extremely costly errors. Further, EMT posits that even if a biased decision-making strategy routinely generates errors, as long as it reliably avoids the errors that are most costly, then the bias should be considered adaptive (i.e., fitness promoting). EMT justifies the book’s rhetorical shift whereby “biases” become “instincts.” Johnson argues that “biases are not decision-making ‘problems.’ Far from it. They are an elegant *solution* to decision-making problems” (24, emphasis in original).

The book’s first pages introduce the main foil for Johnson’s argument regarding the potential benefits of biases: rational choice theory. In *Strategic Instincts*, rational choice theory is defined as “the idea that humans accurately weigh up the

¹ Tobias Gerstenberg and Thomas Icard, “Expectations Affect Physical Causation Judgments,” *Journal of Experimental Psychology: General*, September 12, 2019, DOI: <http://dx.doi.org/10.1037/xge0000670>.

² Martie G. Haselton and David M. Buss, “Error Management Theory: A New Perspective on Biases in Cross-Sex 3514.78. Mind Reading,” *Journal of Personality and Social Psychology* 78:1 (January 2000): 81–91, DOI: <http://dx.doi.org/10.1037/0022-1.81>.

expected costs and benefits of available options and their probabilities of success, and then select the one that is expected to produce the highest utility” (36). Thus, “adaptive decision-making heuristics” are set up in contrast to more effortful calculations driven by utility maximization. *Strategic Instincts* is by no means the first book to consider foreign policy decision-making in light of documented deviations from the predictions of rational choice.³ Nevertheless, *Strategic Instincts* offers a unique take by focusing on the positive outcomes associated with such deviations.

In the rest of the book, Johnson devotes two chapters to each of three biases/instincts: overconfidence, the Fundamental Attribution Error, and ingroup/outgroup bias. For each bias/instinct, Johnson’s core argument is that it can contribute to successful outcomes in the realm of international politics, war and crises in particular. Arguments are spread across two chapters per bias. One chapter is devoted to a discussion of the evidence, largely from psychology and biology, for the bias’s existence and adaptive characteristics. The second chapter considers the bias/instinct in light of a case study. Specifically, Johnson argues that overconfidence aided General George Washington (Chapter 4), that the absence of the FAE hindered Prime Minister Neville Chamberlain (Chapter 6), and that the “right” amount of ingroup bias aided the Americans during the war in the Pacific while excessive degrees of both ingroup and outgroup biases hindered the Japanese (Chapter 8).

If it seems as though this is a lot of ground to cover, it is. If there is a weakness with *Strategic Instincts*, it is that relatively tight definitions of each bias within the corresponding theory chapter give way to conceptual stretching in the case studies. For example, the Fundamental Attribution Error is explored in the context of the perception of Nazi leader Adolf Hitler by British leaders prior to World War II. Within the case, the FAE comes to mean more than the attribution to our foes of dispositional, as opposed to situational, motivations. Rather, the FAE expands to include the notion of threat perception: “Indeed, if the FAE is absent, we run the risk of failing to recognize, prepare for, and deal with threats before it is too late” (171). The book does not make it entirely clear, however, why motive *mis*-attribution should necessarily aid in detecting genuine dangers.

The second question posed at the book’s outset – *when* do cognitive biases contribute to success – has a less obvious answer. Johnson clearly states that context matters in the working of each bias. He also repeatedly notes that, for each of these biases/instincts, moderation is key; extreme levels of these biases are not what is considered adaptive. This is consistent with an important distinction between the study of outliers and the study of species-typical features of the human mind.⁴ But it is sometimes difficult to discern the precise contextual factors and “levels” that bring these psychological mechanisms into play. For example, Johnson uses the case of Washington’s persistence against the British, under unfavorable circumstances, to make the argument that overconfidence can contribute to successful outcomes. The evidence presented in the case study makes Washington sound quite exceptional in the degree of his overconfidence, however: “There is little doubt that this great confidence—a confidence which exceeded that of many of his contemporaries—helped to lift the scattering of poorly prepared volunteers and militiamen, against all odds, to defeat the British Empire” (103). The case study on group biases and the war in the Pacific makes a much stronger argument for the importance of “average” levels of bias. The evidence presented from public opinion surveys, surveys of the armed forces, and accounts of training regimens provide a clearer picture of the biases at work “on average” in large groups (and thus also the range of bias “values” within such groups).

As a final point, the stated goal of *Strategic Instincts* is “to turn the literature on its head” (291). Indeed, the book raises several provocative questions for IR scholars to consider.

³ Robert Jervis, *Perception and Misperception in International Politics* (Princeton.: Princeton University Press, 1976); Rose McDermott, *Risk-Taking in International Politics: Prospect Theory in American Foreign Policy* (Ann Arbor: University of Michigan Press, 2001).

⁴ Anthony C. Lopez and Rose McDermott, “Adaptation, Heritability, and the Emergence of Evolutionary Political Science,” *Political Psychology* 33:3 (2012): 343–362, DOI: <https://doi.org/10.1111/j.1467-9221.2012.00880.x>.

First, should our theories proceed from the assumption that humans can accurately assess the ground-truth of complex situations? *Strategic Instincts* draws on an impressive array of experimental and observational research conducted primarily in the fields of biology and psychology. In this foundational work, biases are often tested relative to a known ground-truth (i.e., pay-off probabilities in a behavioral game). That is, biased perception is defined relative to *accurate* perception. But in the domain of international politics, many ground-truths (e.g., risk levels, motivations) are unknown, even after the fact. Accuracy may be an even more inappropriate benchmark because biases distort fundamental perceptions. These cognitive processes are not blinders (or rose-tinted glasses) that can be removed at will; they are background operations performed by our brains on a continuous stream of sensory input. Assuming we could switch off such processes may be too unreasonable to be a useful analytic move.

Second, how should we study and define success? Johnson notes that the call to study success has a long history within IR scholarship (5).⁵ Within *Strategic Instincts*, success is defined in narrow terms: being on the winning side of a war or a crisis. This approach is analytically tractable, but also incomplete. As Johnson's arguments make clear, the psychological factors that contribute to any day where bad things *don't* happen adds something to the "success" side of the ledger. Identifying the causal mechanisms that contribute to failures is the easier problem, for reasons that Johnson and others outline (see 5-7).⁶ If we define success as simply the absence of failure, then such explanations could serve double-duty. Yet, *Strategic Instincts* suggests that approach misses out on an opportunity to more rigorously identify the psychological factors that contribute actively to our successes.

My own work on threat perception – another cognitive process that is shared across humans, helpful in moderation, and likely to contribute to successful outcomes – grapples with these questions as well.⁷ While *Strategic Instincts* does not contain all of the answers, it provides yet more evidence that these are questions worth confronting. This is especially true for those of us who study how being human – with all the biases and baggage that entails – matters for the ebb and flow of international politics.

⁵ For example, see Robert Jervis, *How Statesmen Think, The Psychology of International Politics* (Princeton: Princeton University Press, 2017), www.jstor.org/stable/j.ctvc775k1.14; Jonathan Mercer, "Rationality and Psychology in International Politics," *International Organization* 59:1 (January 2005): 77–106, <https://doi.org/10.1017/S00208183050058>; Geoffrey Blainey, *The Causes of War*, 3rd ed. (New York: Simon and Schuster, 1988).

⁶ Jervis, *How Statesmen Think*.

⁷ Marika Landau-Wells and Rebecca Saxe, "Political Preferences and Threat Perception: Opportunities for Neuroimaging and Developmental Research," *Current Opinion in Behavioral Sciences* 34 (August 2020): 58–63, DOI: <https://doi.org/10.1016/j.cobeha.2019.12.002>.

REVIEW BY JANICE STEIN, UNIVERSITY OF TORONTO

Dominic Johnson has written an important and provocative book, one that every student of international politics should read. He challenges the conventional wisdom that cognitive biases explain the mistakes that leaders make in the design and execution of strategy and proposes an evolutionary perspective that sees cognitive biases as design features that evolved to improve decision making and strategy. These design features, he argues, protect states against worst-case outcomes.

Johnson develops his argument by demonstrating how three important and ubiquitous biases – overconfidence and the fundamental attribution error, identified by cognitive psychologists, and in-group/out-group bias, identified by social psychologists – worked to empower American revolutionary leaders in their revolt against Britain, helped Winston Churchill, when he was in opposition, to get Nazi leader Adolf Hitler's intentions right, and strengthened American support for the long war in the Pacific. Prime Minister Neville Chamberlain, who never committed the fundamental attribution error, made situational rather than dispositional inferences about Hitler's intent with catastrophic results.

Johnson makes an overarching and important argument about the ways people think. We label patterns of cognition as “biases” only in relation to the standard micro-economic model of rational choice, an imported model that has limited empirical support in the analysis of strategic decisions. Most of us carefully calculate likely costs and benefits only for a few very big and important questions and even on those kinds of questions we do not always meet the standards of rational choice.

Let me go a step further than Johnson does and suggest that we abandon the term “bias” completely and use the terms “cognitive patterns,” and “heuristics,” or cognitive shortcuts. The label “bias” is misleading because it suggests fault or error, precisely what Johnson does not want to convey. What he is telling us is that patterns of cognition – to be overconfident, to infer that somebody’s threatening behavior is intentional, and to stigmatize those who are outside our own group – are universal patterns of thinking that evolved over time as adaptive solutions to some kinds of problems.

First, these patterns of thinking were designed to solve problems that recurred over and over in our evolutionary past. And second, in evolutionary biology, these problems affected the reproduction of the species. It is important to note that it is reproduction, not survival directly, that drives natural selection and adaptive problem solving the individual level.¹ These distinctions matter because it is possible, for example, that individuals or groups may choose to die in order to protect those who carry their genes. They also matter when we transfer concepts from evolutionary biology first to psychology and then to international politics, where the survival of states or groups or leaders has obvious meaning, but reproduction does not.² There are therefore important breaks in the logic chain when Johnson moves from evolutionary biology to evolutionary psychology to international politics that need to be unpacked. Throughout the book, Johnson focuses exclusively on survival of individuals and groups as the evolutionary imperative. The evolutionary imperative, however, is species success, not individual or group survival.³

¹ Charles Darwin, *The Descent of Man* (London: Penguin, 2004); Thomas Dobzhansky, *Mankind Evolving: The Evolution of the Human Species* (New Haven: Yale University Press, 1962).

² The leading research program in evolutionary psychology is John Tooby and Leda Cosmides, The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environment, *Ethology and Sociobiology* 11 (1990): 375-424. Critics argue that their claims are too extensive. See Matteo Mameli, “Evolution and psychology in philosophical perspective,” in Robin Ian MacDonald Dunbar and Louise Barrett, eds. *Oxford Handbook of Evolutionary Psychology* (New York: Oxford University Press, 2012): 21-34.

³ Peter K. Hatemi and Rose McDermott, *Evolution as a Theory for Political Behavior*, 2011. See https://www.researchgate.net/profile/Rose_Mcdermott/publication/260322007_Evolution_as_a_Theory_for_Political_Behavior/links/53dc345f0cf2a76fb667b4d4/Evolution-as-a-Theory-for-Political-Behavior.pdf accessed 25 January 2021. Leda Cosmides and John Tooby, Evolutionary Psychology Primer,” Center for Evolutionary Psychology, University of California, Santa Barbara. See

Johnson is otherwise generally careful to qualify his arguments. Cognitive patterns, he tells us, don't work equally well in all settings and they vary from person to person and group to group as well as from situation to situation. Indeed, there is very large variation within these broad cognitive patterns that are typical of species over time. Several important implications follow from this variation. First, we need to understand the sources of variation across individuals and groups. Some variation is explainable as adaptive problem solving, some as unintended by-products of an adaptation, and some simply as noise.⁴ Johnson tends to treat cognitive patterns as universals and consider them all as adaptive problem-solving, with little attention to variation among individuals and to the possibility that some of these patterns may be unintended by-products of adaptation. Not all patterns, in other words, are design features. Some are bugs.

Second, because these patterns vary across persons and across situations, these patterns can be too strong or two weak, and when they are either, Johnson acknowledges, they can be harmful. To be useful, he argues, cognitive patterns must be manifested in appropriate settings and in moderation. In evolutionary psychology, cognitive patterns are appropriate to the adaptive problem over time. In transferring the principle of reproductive success to international politics, we have to be careful *not* to use environmental attributes rather than problems to specify what counts as "appropriate" behavior. Johnson comes very close to doing so, particularly in his analysis of the international environment that some scholars describe as anarchic.

Third, the concept "moderation" is not one that evolutionary psychologists use in any systematic way. It is also very difficult to define the thresholds when we apply it to international politics. Yet it is an important, if largely unspecified, concept in Johnson's analytic toolbox when he walks us through the excessive in-group bias both in the Japanese military and the excessive out-group bias of American leaders during the Pacific War. At best we can say that we know excess when we see it, but we often know it only after the fact through the harms that it produces.

One further point on the backward-looking focus of evolutionary psychology. Cognitive patterns that evolved over time because they provided adaptive solutions in the past do not necessarily generate adaptive behavior in the present. Johnson acknowledges as much when he argues that these strategic instincts, or amalgams of cognitive patterns, are only likely to confer advantages today if the situation is sufficiently analogous to those in which they evolved. It hinges, as he says, on the matches of patterns that evolved over time from hunter-gatherer societies to our own times. When there is a poor match, evolved patterns of thinking can be ill-suited to the adaptive problems that grow out of the current situation and cause harm.

This is especially challenging when evolved patterns of thinking are applied to international politics. There are similarities in adaptive problems in different environments – first families, then clans, then tribes, then city states, and then the large and complex states and international institutions enmeshed in the patterns of governance that we know today. But there are also important differences. The differences become especially important where quick responses to adaptive problems engage with complex technological systems that are difficult to reverse once they are set in motion. Johnson analyzes the importance of overconfidence in sustaining American rebel forces against the British, but Rose McDermott warns about the impact of overconfident decision makers who have to make rapid-fire decisions about deployment of nuclear weapons under conditions of strategic instability.⁵ It is difficult to argue that cognitive patterns that evolved over time will be well suited to

<https://www.cep.ucsb.edu/primer.html>, accessed January 25, 2021. They argue that a "phenotypic design feature *causes its own spread through a population* (which can happen even in cases where this leads to the extinction of the species). Emphasis in original

⁴ Hatemi and McDermott, 2011.

⁵ McDermott, "The Influence of Psychological Factors in the Search for Strategic Stability," Paper presented at the Center for International Studies, Stanford University, 2020; Dominic D.P. Johnson, Rose McDermott, Emily S. Barrett, Jonathan Cowden, Richard Wrangham, Matthew H. McIntyre, and Stephen Peter Rosen, "Overconfidence in Wargames: Experimental Evidence on Expectations, Aggression, Gender and Testosterone." *Proceedings of the Royal Society B: Biological Sciences* 273, 1600 (2006): 2513-2520.

adaptive problem solving when complex systems of highly lethal weapons that can inflict mass destruction shape security environments.

Johnson then goes beyond evolutionary psychology to situate cognitive patterns and heuristics within the broader context of the challenges of decision making under uncertainty and asymmetric costs. Johnson uses error management theory to think about balancing different types of errors in decision making on international issues.⁶ Robert Jervis characterizes these errors as excessive vigilance, which generate false positives, or excessive complacency, that generates false negatives.⁷ Error management theory concentrates on three kinds of challenges – dangerous objects or people, perception of others, and perception of self – as fundamental to adaptive problem solving. Cognitive heuristics, Johnson claims, are effective in balancing between these two classic kinds of errors of false negatives and false positives.

Error management theory, ironically, relies on at least some rough calculation of probability – how likely is it that this person or state will threaten me? – and some rough calculation of cost and benefit – how dangerous is this state? What damage can it inflict? And then decision makers have to compare across these rough estimates of likely costs and benefits to determine which is the less likely and damaging error.

It is not clear how cognitive patterns are *better* at error management or how, as Johnson argues, the need to balance errors can demand a bias rather than accuracy. For that to be true, Johnson's claim that bias works well *on average* to manage error would have to be validated. But how can we validate the claim? In international politics where cases are, relatively speaking, so few, it is problematic to think in terms of averages.⁸

Johnson argues frequently throughout the book that it is far better to run the risk of false positives when inferring threatening intentions than it is to underestimate threat – a false negative – and risk exploitation and at the outer edge, destruction. His argument does not rest on a law of averages. Rather, it makes clear that even one false negative – Chamberlain's situational estimate of Hitler's intentions – put Britain at risk of a catastrophic defeat. Implicit in Johnson's argument is that avoiding that one false negative is worth any number of false positives and it is in that sense that the fundamental attribution error is an adaptive strategic instinct to manage uncertainty and the asymmetric costs of false negatives and false positives.

Is that argument correct under all conditions? To imagine a hypothetical, a government that saw threats as intentional, armed itself to the outer limits of its capacity, and preemptively attacked at the first opportunity would likely sooner or later exhaust its material and human reserves and make itself vulnerable to an opportunity-seeking antagonist that also saw threatening behavior as intentional. And antagonists aplenty there would be, as other states inferred threatening intent

⁶ Martie G. Haselton and David M. Buss, "Error Management Theory: A New Perspective on Biases in Cross-Sex Mind Reading," *Journal of Personality and Social Psychology* 78:1 (2000): 81-91; Randolph M. Nesse and George C. Williams, "Evolution and the origins of disease," *Scientific American* 11 (1998): 86-93; Gerd Gigerenzer and Henry Brighton, "Homo Heuristics: Why Biased Minds Make Better Inferences," *Topics in Cognitive Science* 1 (2009): 107-143.

⁷ Robert Jervis, "Cooperation under the Security Dilemma," *World Politics* 30:2 (1978): 167-214.

⁸ Johnson identifies Chamberlain's strategy of appeasing Hitler, as well as Soviet leader Joseph Stalin's surprise at Hitler's attack in 1941, and Israel's surprise at the joint attack by Egypt and Syria in October 1973 as three examples where the failure of leaders to make dispositional attributions led states to make themselves vulnerable to unexpected attack (132-133). The interpretation of Israel's leaders is incorrect. Israel's Military Intelligence assumed an intention by both Egypt and Syria to attack (thus making dispositional attributions) but argued that their air forces were still too weak to challenge Israel's superiority in the air and that Egypt and Syria would not attack until they had compensated for that deficit. Paradoxically, they got the intention right, but the estimate of the importance of air power wrong. The other two cases deal with Hitler's Germany, where leaders made situational interpretations and were terribly wrong. Johnson identifies other cases where dispositional attributions led to dangerous escalation, such as the Cuban Missile Crisis. These cases demonstrate the variation we spoke of earlier. It is unclear what the universe of cases is and how the average is calculated.

from the growing military capabilities and aggressive behavior. Thomas Hobbes's world of "the war of all against all" advantages the largest and the richest that can sustain the cost of struggle the longest.⁹

Of course, Johnson argues for moderation. Cognitive patterns taken to extremes are dangerous. But how exactly can leaders practice moderation in assessing threatening intentions when they are biased to explaining threatening behavior as intentional rather than as a response to situational constraints? Layer on top of that an embedded pattern of thinking that foregrounds the asymmetrical costs of even one error, and moderation becomes difficult in theory and in practice, as do the concepts of error management and balancing risks. Even for the largest and the richest states, consistent attribution of threatening intent with no regard for false positives is likely a self-defeating strategy over time. If this argument is correct, the evolutionary advantages of the fundamental attribution error at a time when the costs of war are very high are questionable.

We are left with a puzzle. Johnson rightly asks why humans think in the systematic ways they do. He does not ask how we explain the large variations within the universal patterns that he identifies. Evolutionary theory would suggest that some of our most fundamental ways of thinking are indeed design features to solve adaptive problems and that it is not helpful to think of them as biases. But some are bugs, and we need to be able to identify those. The important questions then become which of these design features match the adaptive problems we face today, and which do not. And for those that do not, how do we build in decision protocols and constraints to reduce the harm of these patterns as we seek solutions to the adaptive problems of the world we live in today.

⁹ Thomas Hobbes, *The Matter, Forme and Power of a Commonwealth Ecclesiastical and Civil*, commonly referred to as *Leviathan*, published in 1651.

RESPONSE BY DOMINIC D. P. JOHNSON, UNIVERSITY OF OXFORD

It is a great honor to have four such esteemed colleagues, Valerie Hudson, Joshua D. Kertzer, Marika Landau-Wells, and Janice Stein, debate my book *Strategic Instincts*, and I am immensely grateful for their insights. Books are such a massive personal and professional investment that anyone taking the time to read them carefully afterwards, let alone put pen to paper, is a rich reward indeed. As well as recognizing the book's goals as new and interesting, they also identify some important critiques, questions, and areas for further study. In my response, I focus on the problems and opportunities they identify, especially those on which two or all the reviewers converged.

Does Individual Variation Undermine the Explanatory Power of Cognitive Biases?

Kertzer and Landau-Wells both highlight a puzzle: cognitive biases are presented as universal aspects of human nature, yet the book's cases often focus on exceptional individuals (such as U.S. General George Washington and British Prime Minister Neville Chamberlain). This raises the question, Might the same outcomes be better explained by personality differences rather than by universal cognitive biases? More generally, if there are individual differences in the expression of cognitive biases, (or the former swamp the latter), how useful are cognitive biases for making predictions about systematic outcomes?

This is an important point. To make the same argument, the book could have focused solely on individuals (and their biases), or focused on the broader populace (and thus biases on average). I included a mixture of both. This was done in order to concentrate the main parts of the analysis on key individuals making the decisions (like other work in political psychology), since this allows us to most clearly identify any hypothesized causal chain between psychological phenomena and state behavior. Where possible, however, I also contrasted these leaders with other people in decision-making positions and the wider military or public. That contrast can itself be a revealing and important source of within-case variation in the strength of bias and the size of effects. But when focusing on individual leaders (or even a group of them), how do we know: (a) that cognitive biases were responsible for their judgments and decisions (as opposed to any number of other person-specific traits); (b) whether these cognitive biases would play (or would have played) the same or as strong a role in a different leader; or (c) whether the bias and its effects generalize to other people down the hierarchy of political decision-makers, advisors, bureaucrats, soldiers, and citizens.

These are important questions both theoretically and empirically. Very briefly on the empirical side, this is at least something that can be tested. To what extent was overconfidence, for example, unique to Washington, common among other leaders, or shared more widely in the population? Landau-Wells contrasts the exceptional characteristics I highlight of Washington, versus the common characteristics I generalized across many actors in the Pacific case. That is an interesting observation. In both cases I made efforts to assess whether levels of bias among leaders was stronger or weaker than those around them, or among third parties. Whether a given bias is widespread or limited to individuals would be worthy of further exploration. But the bigger question at hand is *whether* cognitive biases can have effects on politics. If there is variation (for whatever reason), then political outcomes are most likely to be swayed by biases when they are exhibited by those in power—however strong they are and however much they differ from those around them. So, I think that the focus on leaders is essential for causal inference and impact on political outcomes.

As for the theory, I see individual variation as very much *part* of the story, not an alternative story. Certainly, cognitive biases and personality differences are not mutually exclusive explanations for behavior. Indeed, they interact in important ways. Cognitive biases are often seen as universal—something we all have, everywhere, all the time. However, as with any psychological characteristic, there is in fact considerable variation arising from individual differences and circumstances that alter their expression. Ultimately, this means that we should expect some *distribution* of cognitive biases among the

population as a whole—like a bell curve.³³ The average is some positive manifestation of the bias (most people are somewhat biased much of the time), but there is variation around the mean.

But what are the sources of this variation? There are at least three broad categories of variation: (1) individual variation, (2) contextual variation, and (3) cultural variation. All of these are not only commensurate with psychological and evolutionary theory, but are actually predicted by them. At the most basic level, *individuals* vary in cognitive traits because of genetic differences (our genetic makeup varies through replication and recombination, leading to different levels or patterns in the manifestation of most traits). As with other animals, humans differ in the expression of physiological and psychological traits, even if the broad set of traits they have is species-typical.³⁴ Interactions with other individual characteristics can also generate differences. Variation also arises from context. Temporal or contextual activation of a given trait (for example, we can become angry, but we are not angry all the time). In other words, the strength of a given bias can depend on when and under what conditions one looks.³⁵ There are also important selection effects, such that certain positions and professions may become populated by people with some kinds (or levels) of traits more than others. Finally, there are of course also cultural differences in the expression and manifestation (if not the presence) of cognitive biases, and/or of mediating factors.³⁶

In short, I agree individual variation is important. But that is no different from most physiological and psychological traits. Biological traits are variable and have some kind of naturally varying distribution. In my mind, I imagine many of the leaders discussed as representing the tail ends of the distribution of cognitive biases (imagine a bell curve of, say, overconfidence in the population at large, with Washington being on the right-hand side of this curve). Psychologists have argued that, for a given bias, there is likely to be some “optimal” level for it to function effectively.³⁷ Other authors have in fact explicitly argued that leaders are likely to represent *extremes* of psychological traits, rather than being less biased or typical.³⁸ Whether leaders with high-levels of biases gravitate to the job, tend to be selected by others in the process of getting there, or find that these traits become exaggerated once in office, whatever biases leaders have can then exert an important impact on state behavior.³⁹ The thing about cognitive biases is that they represent general tendencies, but tendencies that represent an

³³ Roy F. Baumeister, “The Optimal Margin of Illusion,” *Journal of Social and Clinical Psychology* 8 (1989): 176-189.

³⁴ See, for example, Donald E. Brown, *Human Universals* (New York: McGraw-Hill, 1991); Tiffany A. Ito and John T. Cacioppo, “Variations on a Human Universal: Individual Differences in Positivity Offset and Negativity Bias,” *Cognition and Emotion* 19 (2005): 1-26.

³⁵ See, for example, Shelley E. Taylor and Peter M. Gollwitzer, “The Effects of Mindset on Positive Illusions,” *Journal of Personality and Social Psychology Bulletin* 69 (1995): 213-226.

³⁶ See, for example, Incheol Choi and Richard E. Nisbett, “Situational Salience and Cultural Differences in the Correspondence Bias and Actor-Observer Bias,” *Personality and Social Psychology Bulletin* 24:9 (1998): 949-960.

³⁷ Baumeister, “The Optimal Margin of Illusion.”

³⁸ Arnold M. Ludwig, *King of the Mountain: The Nature of Political Leadership* (Lexington: University Press of Kentucky, 2002); S. Nassir Ghaemi, *A First-Rate Madness: Uncovering the Links between Leadership and Mental Illness* (New York: Penguin Press, 2011).

³⁹ In discussing *individual* variation, we should also think about *group* decision-making. How do (or can) cognitive biases at the individual level affect decision-making at the level of the group? Sometimes, different biases acting at each level might cancel each other out, other times they may exacerbate each other (overconfidence is argued to be one example of the latter). See Irving L. Janis, *Victims of Groupthink: Psychological Studies of Policy Decisions and Fiascoes* (Boston: Houghton Mifflin, 1972); Daniel J. Goleman, “What Is Negative About Positive Illusions? When Benefits for the Individual Harm the Collective,” *Journal of Social and Clinical Psychology* 8:2

independent variable (for example, levels of overconfidence), which by definition we *need* to be variable, if they are to explain variation in outcomes. Kenneth Waltz worried that, if human nature is fixed, then it cannot explain variation in war and peace.⁴⁰ Fortunately he was wrong about that. Human nature is not fixed. It has many general tendencies, but these vary in degree from person to person, and so do cognitive biases.⁴¹

Clearly, there are many potential sources of variation in cognitive biases, and we should *expect* them to vary among people rather than be surprised. But then the question becomes, what are the implications of this variation for the argument for *Strategic Instincts*? Does variation mean cognitive biases are more or less likely to lead to *useful* or *detrimental* outcomes? Or, are they more likely to be expressed in favorable or unfavorable conditions? These are interesting questions to explore, and highlight the fact that it may matter a great deal who is making the decisions. The extent of their cognitive biases—variable as they are—may determine whether they lead us into success or disaster.

How Do We Get From Evolution to International Relations?

Perhaps the next most important point to address is the role of evolution in human behavior and how this plays out in international relations. Stein identifies “breaks in the logic chain” in moving “from evolutionary biology to evolutionary psychology to international politics that need to be unpacked.” I agree about the dangers of this extension, and attempted to address this explicitly in the book (in chapter 3), but hopefully can clarify further here. In particular, Stein highlights the important distinction between *survival* versus *reproduction*—technically speaking natural selection only cares about the latter.⁴² This is important because, as she points out, while we have analogies of survival in modern day interactions among humans and nation-states, we don’t have an easy analogy for reproduction. States may compete for survival, but surviving states don’t reproduce more of themselves. Therefore, it is often assumed, evolution has limited applicability to international relations. However, that is not, in fact, the argument here. I am not claiming that evolution is taking place *now* as a process of competition and replacement among states (evolutionary selection).⁴³ Instead, we are claiming that the

(1989): 190-197. But context can matter too. If biases tend to kick in under similar environmental conditions, then even very different people are likely to become biased in the same direction (if not all to the same degree).

⁴⁰ Kenneth N. Waltz, *Man, the State and War: A Theoretical Analysis* (New York: Columbia University Press, 1959).

⁴¹ See, for example, Irene Scopelliti et al., “Individual Differences in Correspondence Bias: Measurement, Consequences, and Correction of Biased Interpersonal Attributions,” *Management Science* 64.4 (2018): 1879-1910. See also Brian C. Rathbun, “The Rarity of Realpolitik: What Bismarck’s Rationality Reveals About International Politics,” *International Security* 43:1 (2018): 7-55.

⁴² This is actually slightly more complicated than people often make out. Natural selection has obviously led to the development of billions of amazing adaptations for survival across the animal kingdom (from armor to plant defenses to poisons). If it only cares about reproduction, why does it bother with mechanisms that aid survival? Because reproduction is only possible if organisms *survive to reproductive age*. Therefore, survival matters too, and traits that affect survival rates are obviously subject to selection pressure. Evolutionary biologists have showed this in action: for example, individuals in more predator rich environments evolve better defenses. The process by which this occurs is of course differential reproduction of those with better defenses, but the outcome is the evolution of traits for survival. See Joel G. Kingsolver and David W. Pfennig, “Patterns and Power of Phenotypic Selection in Nature,” *BioScience* 57:7 (2007): 561-572; N.B. Davies, J.R. Krebs, and S.A. West, *An Introduction to Behavioural Ecology* (Chichester: Wiley Blackwell, 2012).

⁴³ Again, this is actually more complicated. First, from a genetic perspective, individuals come and go but what sticks around are genes and the traits they represent. Some aspects of international relations are like this. For example, states may stay roughly the same but the *traits* they have (such as democracy or cultural norms) can be more or less successful—and spread among other states. So, in some ways, a process of evolution is taking place all the time, even at the level of the international system (albeit as a form of “cultural evolution”). See Alex Mesoudi, *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (Chicago: University of Chicago Press, 2011). Second, while rare, states do sometimes die. And new ones form. More often, over history, states have expanded or contracted rather than disappeared altogether, or aggregated with other states, or become separated. Nevertheless, some of the processes and selection mechanisms that generate these changes can be analogous to the replication of states or

result of evolution over past millennia has shaped the way the human brain works in making judgments and decision-making (evolutionary psychology).

This is important, because it means that there is no prior expectation that evolved traits should be advantageous or disadvantageous today (“adaptive” or “maladaptive,” in evolutionary terminology). That remains an empirical question. Cognitive biases were “designed” for a different environment and era. They might remain useful in conditions where the problems we face today emulate those of our past (and thus our evolved propensities work well in the modern environment). But then again they might not, where the problems we face today are very different (and our evolved propensities lead us to do something detrimental or dangerous). The book sought to do something new in international relations by exploring the former—evolved biases that may remain “adaptive” today—instead of the (more typically studied) latter, which is a case of evolutionary “mismatch” (or “bugs,” as Stein interestingly calls them). In the book section “What does adaptive mean for state behavior” (40), I argued that a given trait could be considered adaptive for a state as a whole, or for a leader’s own goals, based on its positive or negative effects. But either way that is just asking about the contemporary consequences of our evolutionary legacy.

As with other approaches in political psychology, the broader argument is that individual human beings—and their perceptions and behavior—matter for the decisions and actions of a state. We are thus studying *neither* the survival *nor* the reproduction of states. Instead, we are studying the perceptions and behavior of the people within them—whether citizens, soldiers, or leaders. Evolutionary psychology (as well as, not instead of, social and experimental psychology) helps to understand systematic variation in the traits of those human beings. Evolution brings added value because it accounts for the legacy of natural selection that took place *in the past*, over our evolutionary history, not any evolutionary process that is going on today. If you like, evolution gives us a manual to explain why humans are wired the way they are, why they can be expected to do this or that under different conditions. It does not try and explain the international system or the behavior of abstract states per se. It just describes how people are likely to think and behave within and among them.

Thus, evolutionary psychology is not dissimilar from other psychological theories. It says people do X, and X makes states do Y instead of Z. The evolutionary approach just offers an account of why it is we do X, and in so doing highlights why X might be adaptive, or at least why it was adaptive in the past even if it no longer is today. There are no claims that make the absence of “reproduction” in international relations a problem. X is the *result* of past human reproduction and selection. It does not imply or claim anything about reproduction among states or even of humans today.

The Fundamental Attribution Error

Kertzer and Landau-Wells also both picked up on another puzzle in the book—the relationship between *attribution bias* and *threat perception*—and the potential danger of “conceptual stretching” of the former into the latter. This was interesting because I had wrangled with this exact problem in the writing process (especially given the reverse case study on this topic—an *absence* of bias and a *failure*, rather than presence and a success). This is why I was eager to clarify at the start of Chapter 5 on the fundamental attribution error (FAE) that: “This does not mean that we always perceive others’ behavior as threatening but rather that we will perceive threatening behavior as *intentional*” (117). As Kertzer and Landau-Wells note, disposition and threat represent a very important distinction. The FAE means that people overweight dispositional causes of others’ behavior, but the behavior itself could be anything—it doesn’t have to be threatening. There is certainly no prediction of a direct effect in which FAE leads to threat inflation.

state entities. These changes can be human driven or accidental, but other times they are subject to selection effects depending on what works. Peter Turchin et al., “War, Space, and the Evolution of Old World Complex Societies,” *Proceedings of the National Academy of Sciences (PNAS)* 110:41 (2013): 16384-16389.

However, there *is* a prediction that, if states have a prior to be suspicious of other states, and thus focus more on negative actions and information than positive actions and information, *then* the FAE may indeed lead to the exaggeration of threats (more than opportunities) among the events and states they are paying attention to. Do states have such a prior? Yes, for three reasons. First, even rational choice approaches suggests states should (or at least do, empirically) err on the side of caution, lest they be exploited, which is a worse scenario than making the opposite mistake (like in the prisoner's dilemma, or security dilemma).⁴⁴ Second, on top of that, even if good or bad information (or outcomes) occur with equal frequency and magnitude, the “negativity bias” means we tend to *overweight* bad events and information compared to good events and information.⁴⁵ Third, the FAE carries an important reverse prediction: while we tend to ascribe others’ behavior as dispositional, and are even *more* likely to ascribe *bad* behavior by others as dispositional (i.e., on purpose). We ascribe *good* behavior by others as situational (i.e., an accident). Therefore, actions seen as bad—e.g., threatening—are more likely to be seen as dispositional than other kinds of behaviors (such as another state offering an olive branch)—when in fact we become suspicious, and ask, “that’s odd, why are they doing that?”). As Kertzer notes, there are clearly important countervailing and interacting biases, which can be contradictory or complementary. These certainly make things interesting (as well as more complicated), and have important implications for generating differences in perceptions—and perhaps threat perceptions in particular.⁴⁶ They also generate interesting predictions, since we can hypothesize, and study, when and what interactions are likely to apply in different scenarios and lead to differential outcomes.

In short, it is hard to disentangle attribution and threat, but we should expect interactions rather than independence, and the former appears to influence the latter in important ways—especially when other states undertake *negative* actions that are (or we see as) menacing or threatening. Under those circumstances, the balance of risks, as well as other cognitive tendencies, mean that we are especially likely to perceive threats as intentional.

Finally, Landau-Wells notes that, whatever the debates above, it is not clear “why motive *mis*-attribution should necessarily aid in detecting genuine dangers.” As other authors have stressed, a threat is a threat regardless of whether it originates from a foe’s disposition or from their situation.⁴⁷ Either way, they pose a threat that must be dealt with. However, there are reasons why “*mis*-attribution” (or dispositional over-attribution) may be an advantage. Threats that arise from State X’s disposition can pose a trickier problem than those that arise from situational factors, because it means X may be harder to deter or dissuade. Their goals and ambitions are resistant to change (if it is situational, we can rectify or change the

⁴⁴ Robert Axelrod, *The Evolution of Cooperation* (London: Penguin, 1984).

⁴⁵ Roy F. Baumeister et al., “Bad Is Stronger Than Good,” *Review of General Psychology* 5:4 (2001): 323-370. Moreover, people have been shown to be more likely to attribute agency—of any kind at all—to negative than positive events. Carey K. Morewedge, “Negativity Bias in Attribution of External Agency,” *Journal of Experimental Psychology* 138:4 (2009): 535-545.

⁴⁶ In other work, Dominic Tierney and I have explored in particular how we reconcile positive biases (e.g., optimism) with the apparent opposite, negativity bias. See Dominic D.P. Johnson and D.R. Tierney, “Bad World: The Negativity Bias in International Relations,” *International Security* 43:3 (2019): 96-140. Both are widespread and powerful. However, part of the variation can be explained by the subject of the bias—perceptions of self lead to positive biases, and perceptions of others (or the environment) lead to negative biases. This is interesting because (a) it means biases can kick in to offer different adaptive advantages in different kinds of situations (there are different biases for different kinds of problems); and (b) there can be complementary or toxic combinations. Most notably, we point out that negative biases to see a dangerous world, combined with overconfidence about one’s ability to succeed within it, is a recipe for disaster. These kinds of interactions among biases suggest a lot of interesting consequences for strategic instincts too—when biases are good may depend not only on the situation, but also on what other biases may be at work.

⁴⁷ Most notably, of course, a Waltzian realist would argue that all state behavior has a situational cause in the anarchic international system, and threats are therefore the result of states being forced into action because of their environment, not their intentions *per se*. Kenneth N. Waltz, *Theory of International Politics* (New York: McGraw-Hill, 1979).

situation; if it is dispositional, we cannot so easily change, or even observe, their disposition—they are sticky by definition).⁴⁸ This was of course the case with Adolf Hitler in 1930s Germany. Chamberlain did change the situation (bowing to reunification of the German speaking peoples, which was the limit of Hitler's stated situational problem), but that did nothing to change his broader dispositional aspirations. Thus, dispositional threats may actually be more dangerous than situational threats. If so, it is good to be more alert to the former—and more than we might think.

Too Hard on Political Psychology?

Kertzer suggests that the book was rather harsh on political psychologists. That's an important charge and one I should qualify. It is risky to criticize, colleagues for invoking biases as causes of mistakes in international relations. For one, these same colleagues have done the hard work of bringing the cognitive revolution to the attention of the academic and policy communities in the first place. And second, many cognitive biases *do* indeed cause mistakes and disasters. I've spent many years arguing that cognitive biases are a source of policy mistakes myself, so I certainly don't intend to upend that approach.⁴⁹ I just question whether it is the whole story.

Perhaps more importantly, while political psychologists have many differences in methods, approaches, and opinions, there are much bigger intellectual opponents to argue with—such as blank-slatists, hard rational choicers, or structural realists, who deny any important role for cognitive biases (and often even psychology or individuals) at all. They were targets in the book too (see, for example, Chapter 2).

However, there does remain a serious question about the stance of political psychology with regard to cognitive biases. Given that the point of the book was to contrast “biases as bad” with “biases as good” (or rather, “heuristics as good”), a necessary critique was why “we” as a subfield tend to assume that cognitive biases are always detrimental, and why we don’t look at successes. In the big scheme of things, Step 1 for the field was to bring the cognitive revolution into IR. That has been done pretty successfully. And the way to do it effectively was to show that: (a) empirically, cognitive biases mean behavior deviates from rational choice and other theories’ predictions; and (b) this is important because cognitive biases can cause serious policy failures, disasters, and wars. If that is the case, then we need to take them seriously.

But Step 2 should perhaps now be to broaden out, to look at the effects of biases, whatever they are—positive as well as negative. This was the point of the book and the point of departure from “normal” political psychology. As the Nobel laureate Daniel Kahneman says, we express these biases all the time in everyday life, and we seem to get along OK.⁵⁰ Indeed, dealing with many of life’s demands would be impossible without them. For all our foibles, human beings are pretty remarkable organisms and even our political leaders sometimes achieve amazing things, or at least succeed in persevering to minimize disaster in hard times. Probably, being human with all of our cognitive biases helps.

A Step 3 remains to be done as well. If biases can help us as well as hurt us, we need to think about a broader theoretical framework for human cognitive biases—*why* do they help? They are not just a random collection of conflicting idiosyncrasies, or a systematic set of processing limitation or errors. Rather, they are adaptations. As part of the make-up of the human mind, an evolutionary approach to cognitive biases and their associated behaviors gives us the tools to develop a

⁴⁸ Situational causes can in principle be observed, measured, and altered; dispositional causes are, by contrast, hard to discern and understand let alone change, as stressed by Hans J. Morgenthau, *Politics among Nations* (New York: McGraw-Hill, 1993 [1948]).

⁴⁹ For example, in Dominic D.P. Johnson and Dominic R. Tierney, *Failing to Win: Perceptions of Victory and Defeat in International Politics* (Cambridge: Harvard University Press, 2006); Dominic D.P. Johnson, *Overconfidence and War: The Havoc and Glory of Positive Illusions* (Cambridge: Harvard University Press, 2004).

⁵⁰ As he put it, “Our thoughts and actions are routinely guided by System 1 [our intuitive thinking] and generally are on the mark.” Daniel Kahneman, *Thinking, Fast and Slow* (London: Allen Lane, 2011). 416.

more grounded theory for their origins, causes, and consequences. What were they originally intended to achieve? And thus under what conditions are they more or less likely to continue to bring adaptive advantages today (and when not)—to individuals, groups, and states?

All three reviewers, I was delighted to see, pick up on the highly contingent nature of strategic instincts. Are they good or bad? It depends. The same bias could be disastrous for one leader or in one situation, and clinch victory in another (Napoléon Bonaparte's confidence led him to great victories as well as his Waterloo; Washington's confidence led him into defeat after defeat, but ultimately dragged him to victory). This context dependence is crucial, and has many sources of variation: which bias are we talking about? What context? Among whom? Who is the “recipient” of the bias's changed behavior? Where does it occur in the hierarchy of decision-making and action? At what point in an interaction does it arise? What does it signal to domestic or foreign audiences? All of these things matter to whether a bias will: (a) be manifested strongly, or at all; and (b) whether it will lead to good or bad outcomes. But all this variation is good for business. Since cognitive biases have systematic psychological and physiological triggers, we can derive predictions for *when* they are likely to help and *when* they are likely to lead us astray. Many opportunities to explore and chart this taxonomy await.

What Should we Do About Biases?

Landau-Wells nicely highlights how cognitive biases should not be viewed as unusual or exceptional phenomena, but as constant “background operations” acting on all of our perceptions, and thus something we cannot just “switch off.” This raises the question of what we should (or even can) do about them. As Stein notes, while some might be useful (at least sometimes), many human cognitive tendencies remain important “bugs” that are liable to lead us into error or disaster, and it is vital to “build in decision protocols and constraints to reduce the harm of those patterns.” While I argue in the book that biases can be adaptive, I was certainly not arguing that all biases are advantageous all the time. Many of them remain harmful, and some that are useful in one context can be harmful in another. In the book I noted that even Kahneman says that, despite having spent his entire career identifying and studying biases, it has not helped him avoid them himself. Asked in an interview about his retrospective book, *Thinking Fast and Slow*, he replied, “I don’t think reading this book will help you. Writing it certainly hasn’t helped me!”⁵¹ I therefore concur that rather than more information and attempts at anti-bias training, we need to change the structure, or hardware, within which people make their (biased) judgments and decision-making, since we cannot reliably update their software. In particular, situations of high-stakes decisions, crisis, and war, are precisely the times when people do *not* have the luxury to consider all options and think through their various consequences and ramifications. This is when biases come to the fore and mistakes become more likely and more severe. *Strategic Instincts*, however, also suggests a caveat: sometimes, it may be that cognitive biases *help* us navigate those tough times, and steer us towards making good, rapid, gut decisions, when rational decision-making would fail, or hesitate to make any decision at all. And of course even rational decisions can lead to disaster too.

Might Cognitive Biases Lead to Other Kinds of Successes (Beyond War)?

Landau-Wells also points out that the book focused on one particular type of success—success in wars or crises. But what about other types of success? What are the consequences of cognitive biases for international cooperation, conflict resolution, or the formation of institutions? Maybe it helps those endeavors as well as sometimes undermining them. This is a great and important point and opens up many avenues for new work. Now that we have debated the role of biases leading to failures and successes in conflictual interactions, what about the good side of human nature and its effects on many other areas and issues in international relations? I did suggest in the book that many of our biases are likely to help make good things happen, such as—under the right conditions—peace efforts, alliances, cooperation, and moral norms. Indeed, they may compel some people to commit their lives and livelihoods to the pursuit of justice, peace, and cooperation—which are often seemingly beyond the point that an unbiased rational actor would be willing to commit their time and

⁵¹ Christian Jarrett, “A Journey in the Fast and Slow Lanes,” *The Psychologist* 25:1 (2012): 14-15, 15.

effort, or long after they would have given up. In fact, there is a vast literature that focuses on the cognitive adaptations underlying many positive human traits, such as cooperation,⁵² collective action,⁵³ and moral behavior.⁵⁴ Many of these appear to deep-seated evolved adaptations that were important in our evolutionary history and continue to help us thrive today. Some of them go back even beyond our own species: one of key ideas in the biological anthropology of the great apes—of which we are one—is the important role of coalition building, peace brokering, and peacekeeping in social group dynamics, which has been documented extensively, for example, in chimpanzees.⁵⁵ Evolution has endowed us with many prosocial traits for helping and reputation building as well as competing and fighting.⁵⁶

Be warned, however, that turning our attention to positive behaviors is not a panacea. First, many positive behaviors are ways of dealing with the negative—coalitionary adaptations to manage the ever-present threat of tyranny, cooperative adaptations to navigate the ever-present threat of self-interest and exploitation, and moral norms to promote in-group cohesion in the case of out-group competition. Second, whether positive or negative, the advantages of evolved traits can always be broken because of the mismatch between the past and current environment. Even positive biases to help others and cooperate can lead us into danger today if they leave us vulnerable or open to exploitation. So just as conflictual cognitive biases can sometimes be advantageous (as I argue in the book), cooperative biases can also sometimes be detrimental rather than good. There is nothing more dangerous than a tightly cohered group with powerful moral norms they want to spread.

Conclusion

At the end of the day, *Strategic Instincts* proposes a very simple idea. Cognitive biases are nearly always seen as detrimental—especially in the world of international politics. The human brain often seems to be another liability we have to worry about that gets in the way of good policy making and maintaining the peace. But we have these biases in the first place because they helped us make good decisions, not bad ones. Might, therefore, these same cognitive biases sometimes still lend an advantage today, even in the complex domain of international relations? My answer is, yes, they can. Perhaps especially in times of crisis and war, when we need all the motivation and perseverance we can get, they promote action instead of inaction, efficiency instead of perfection. As to exactly when, where, and to what extent biases help rather than hurt, especially given their often simultaneous positive and negative effects and the work of other interacting biases, much remains to explore. This book was a first cut at thinking about possible examples and consequences of cognitive biases as “strategic instincts,” rather than stupid mistakes. Hopefully I have shone a torch in a new direction worthy of further exploration, and one that was only really illuminated by taking an evolutionary perspective on human behavior.

⁵² Matt Ridley, *The Origins of Virtue: Human Instincts and the Origins of Cooperation* (London: Penguin, 1996).

⁵³ Ernst Fehr and Urs Fischbacher, “The Nature of Human Altruism,” *Nature* 425 (2003): 785 - 791.

⁵⁴ Oliver Scott Curry, Daniel Austin Mullins, and Harvey Whitehouse, “Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies,” *Current Anthropology* 60:1 (2019): 47-69.

⁵⁵ Frans B. de Waal, *Peacemaking among Primates* (Cambridge: Harvard University Press, 1989).

⁵⁶ Manfred Milinski, Dirk Semmann, and Hans-Jurgen Krambeck, “Reputation Helps Solve the ‘Tragedy of the Commons’,” *Nature* 415 (2002): 424-426.