

*Review Article*

REDEFINING THE DEBATE OVER  
REPUTATION AND CREDIBILITY IN  
INTERNATIONAL SECURITY  
Promises and Limits of New Scholarship

By ROBERT JERVIS, KEREN YARHI-MILO, and DON CASLER

ABSTRACT

A wave of recent scholarship has breathed new life into the study of reputation and credibility in international politics. In this review article, the authors welcome this development while offering a framework for evaluating collective progress, a series of related critiques, and a set of suggestions for future research. The article details how the books under review represent an important step toward consensus on the importance of reputation in world politics, elucidating scope conditions for when reputational inferences are likely to be most salient. The authors argue that despite the significant accomplishments of recent studies, the scholarly record remains thin on the psychology of the perceiver and is instead focused on situational factors at the expense of dispositional variables and is rather myopically oriented toward reputation for resolve to the exclusion of other important types. Despite its contributions, the new literature still falls short of a full explanation for how actors draw inferences about reputation. These remaining theoretical challenges demand scholarly attention and suggest a role for psychology in filling some of the gaps.

Mark J. C. Crescenzi. 2018. *Of Friends and Foes: Reputation and Learning in International Politics*. New York, N.Y.: Oxford University Press, 208 pp.

Frank P. Harvey and John Mitton. 2016. *Fighting for Credibility: US Reputation and International Politics*. Toronto, Canada: University of Toronto Press, 300 pp.

Van Jackson. 2016. *Rival Reputations: Coercion and Credibility in US-North Korea Relations*. Cambridge, UK.: Cambridge University Press, 219 pp.

Danielle L. Lupton. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca, N.Y.: Cornell University Press, 264 pp.

Roseanne W. McManus. 2017. *Statements of Resolve: Achieving Coercive Credibility in International Conflict*. Cambridge, UK: Cambridge University Press, 241 pp.

Perhaps a man's character is like a tree, and his reputation like its shadow;  
the shadow is what we think of it; the tree is the real thing.

— *Abraham Lincoln*

Dropping bombs on someone to prove you're willing to drop bombs  
on someone else is just about the worst reason to use force.

— *Barack Obama*

Superpowers don't bluff.

— Obama adviser Tony Blinken\*

## I. INTRODUCTION

A wave of recent scholarship has breathed new life into research on reputation and credibility in international politics, particularly in the subfield of security studies. In this review article, we unpack the key concepts, highlight collective progress on core questions, and offer several directions for future research. We assess common themes, diagram the central questions at stake, and synthesize recent contributions to this vibrant research agenda.

In what we characterize as a third wave of scholarship on the topic of reputation and credibility, we find that recent research largely agrees that the logic of reputation must be evaluated through the eye of the beholder and address scope conditions like when and how reputation matters.<sup>1</sup> International relations scholarship has long debated whether actors can develop reputations as well as whether decision makers use reputation to assess others' likely future behavior, and the latest research on the subject answers both questions in the affirmative. These findings not only allow the broader research program to focus on when reputational inferences may be more or less salient, but they also point out the potential for disaggregating across different types of reputation.

We highlight several issues that we believe deserve more attention from scholars of reputation and credibility. These include weighing methodological trade-offs and clarifying the appropriate unit of analysis, parsing unit-level differences among observers, and deepening our understanding of the perceivers' psychology. We argue that despite the accomplishments of recent studies, scholars could do more to unpack how and why reputational beliefs and inferences can vary across perceivers and how insights from psychology and related fields could not only shed light on the sources of such differences, but also generate predictions about patterns in variation that are likely to emerge.<sup>2</sup> We further

\* Lincoln quote, "Excerpt: 'The Lincolns'," National Public Radio, July 30, 2008; Obama quote, Jeffrey Goldberg, "The Obama Doctrine," *Atlantic*, April 2016; Blinken quote, Adam Entous, "Behind Obama's About-Face on Syria," *Wall Street Journal*, June 15, 2013.

<sup>1</sup>The first wave of scholarship includes works by Schelling 1966; Jervis 1970; Jervis 1976; Snyder and Diesing 1977; and Jervis and Snyder 1991, which are premised on the centrality of reputation in international politics. The second wave, which was a backlash to the first wave, includes works by Hopf 1994; Mercer 1996; and Press 2005, who argue that international actors do not acquire reputations and that observers do not rely on others' past actions when assessing the latter's current credibility. The third wave, to which the books under review belong, also includes work by Yarhi-Milo 2014; Kertzer 2016; Renshon, Dafoe, and Huth 2018; and Yarhi Milo 2018.

<sup>2</sup>For good reviews of the current state of political psychology as applied to IR, see Kertzer and Tingley 2018; Davis and McDermott 2020. Our purpose here is to focus on how psychological variables intersect with the study of credibility and reputation.

urge researchers to address the tension between situational and dispositional variables in the study of reputation. Moreover, given the rather myopic focus in security studies on reputation for resolve or credibility of threats—with a few important exceptions that we note below—we believe that future scholarship should pay greater attention to important cross cutting trade-offs between different types of reputations and how actors and observers analyze them. These remaining challenges demand further theoretical development and empirical investigation in future scholarship. Accordingly, we conclude with suggestions for research designs that mix inductive and deductive approaches, integrate evidence from both the signaler and the perceiver, probe potential differences between elites and the public with regard to reputational attributions, and establish connections between actors' self-images and concern for reputation.

The article proceeds as follows. First, we begin by laying out the research program's core concepts and controversies. Second, we describe major recent advances in the study of reputation and credibility by synthesizing the key contributions of five books and diagramming the broad points of consensus. Third, we offer critiques of the new scholarship, pointing to tensions that must be resolved to continue pushing the broader research agenda forward. Last, we outline ideas for the next wave of scholars to consider in growing our understanding of when, why, and how reputation operates in international politics.

## II. UNPACKING THE KEY CONCEPTS

Many notable works in politics, economics, and sociology are premised on the belief that reputations are powerful. By reputation, we refer to beliefs about an actor's persistent characteristics or tendencies based on that actor's past behavior, which will influence what he or she does in the future; for credibility, we mean the extent to which an actor's statements or implicit commitments are believed.<sup>3</sup> Reputation involves judgments about character, or "type," with the implication that this character is baked into an actor's disposition and will affect the actor's behavior in predictable ways. Reputation can be general and involve broad beliefs about type, or it can be specific and capture more focused beliefs about an actor's propensity to live up to his or her word (also known as "signaling reputation")—especially in the context of promises and threats.<sup>4</sup> States tend not to bluff, for example, not only because they fear developing a reputation for dishonesty,<sup>5</sup> but also because

<sup>3</sup> Dafoe, Renshon, and Huth 2014, 374.

<sup>4</sup> Jervis 1970.

<sup>5</sup> Sartori 2002; Sartori 2005.

their leaders face domestic audience costs for backing away from public threats.<sup>6</sup> This view of reputation as relatively unchanging and influential contrasts with approaches that stress the importance and prevalence of how observers update their beliefs and how actors themselves change. These approaches suggest that experience and interaction, such as when a state with a history of defaulting on debt manages to pay its bills on time<sup>7</sup> or when a relatively unknown country joins an international organization known for cultivating certain qualities in its members, can alter types.<sup>8</sup> It is plausible, for example, to argue that although the Vietnam War may have revealed something about the American type, it also changed it and made the United States less likely to behave similarly in the future, even under similar circumstances.

Reputation can contribute to credibility, which is thought to equal a state's capabilities times its interests times its reputation for resolve.<sup>9</sup> Credibility is usually associated with the possibility of using force, but it can also apply to other foreign policy instruments, such as economic or diplomatic sanctions, as well as to other dimensions of policy, such as competence or consistency; in these cases, the related determinants of credibility may be similar only at the most abstract level. Some scholars argue that credibility has a large situational component; states use a "current calculus" of others' capabilities and interests rather than a review of their past actions when calculating credibility.<sup>10</sup> In this view, reputations rarely form based on previous behavior,<sup>11</sup> or at least do not accrue in such a straightforward manner.<sup>12</sup> Others have pushed back, arguing that reputational inferences play a substantial role in assessments of credibility because states that back down from challenges are more likely to receive subsequent ones.<sup>13</sup> Yet reality may be even more nuanced: a highly resolved state still may back down under very unpromising circumstances, while one with little resolve is likely to fight if the alternative is extinction. A further complication is that while credibility and reputation are distinct concepts, they tend to overlap in policy and everyday parlance—where being credible is virtually synonymous with having a reputation for qualities like honesty, reliability, or prudence. In particular, when policymakers and pundits publicly declare

<sup>6</sup> Fearon 1994; Tomz 2007a; Weeks 2008; Levendusky and Horowitz 2012; Levy et al. 2015; Lin-Greenberg 2019.

<sup>7</sup> Tomz 2007b.

<sup>8</sup> Gray and Hicks 2014.

<sup>9</sup> Mercer 1996; Tang 2005.

<sup>10</sup> Press 2005.

<sup>11</sup> Mercer 1996.

<sup>12</sup> Hopf 1994.

<sup>13</sup> Weisiger and Yarhi-Milo 2015.

that American credibility is on the line, what they often mean is that the United States' reputation for upholding commitments, defending allies, and deterring adversaries, and as a result, its image as a global leader, underwriter of security, and steward of the international commons is somehow under threat.

International relations scholars have not only investigated the extent to which a state's decisions to stand firm or to back down in past crises affect its reputation for resolve, but also have probed the extent to which such reputations inform observers' subsequent choices and policies.<sup>14</sup> On the one hand, claims for the centrality of reputation are linked to the strategy of commitment, and both of these can be traced to Thomas Schelling: "Essentially we tell the Soviets that we have to react here because, if we did not, they would not believe us when we say that we will react there . . . . Our deterrence rests on Soviet expectations."<sup>15</sup> On the other hand, those who cast doubt on the importance of reputation suggest that states' memories are hardly so specific or perfect, in which case, as Frank Harvey and John Mitton note, there would be a limited price for bluffing and statements of commitment would be downgraded if not ignored (pp. 21–22). This raises tension, if not a contradiction, in arguing that some behavior deeply marks the actor's reputation and claiming that the resulting image will be held over a prolonged period of time in the face of varying behavior. If each event significantly affects the actor's reputation, then how does reputation itself matter? And if reputations are very sticky, how are they established and how much freedom does the actor have to act in ways that are discrepant from its reputation without sacrificing it?<sup>16</sup>

All of this suggests that although we can easily define credibility and reputation, we are only beginning to understand the scope conditions, such as decision makers' psychological traits and beliefs about the nature of the international system, that affect their salience.<sup>17</sup> Some states and leaders who suffer limited defeats certainly behave as if they need to act boldly after backing down. For instance, declassified records make it clear that this was the main motive for the strong American reaction when the new communist government of Cambodia seized the merchant ship *Mayaguez* in 1975. The United States used force before fully exploring the possibility of securing the release through negotiations, and indeed, took actions that President Gerald Ford and Secre-

<sup>14</sup> Yarhi-Milo, Kertzer, and Renshon 2018; Zhang 2019.

<sup>15</sup> Schelling 1966, 55–56.

<sup>16</sup> Mercer 2013b.

<sup>17</sup> Yarhi-Milo 2018.

tary of State Henry Kissinger understood could well lead to the crew's death in an unsuccessful attempt to reestablish a reputation for resolve in the wake of the ignominious retreat from Vietnam.<sup>18</sup> Notwithstanding the need for even greater clarity regarding scope conditions, a series of fresh studies has broken the stalemate over whether reputations form at all, and significantly add to our understanding of when they are likely to affect outcomes in international politics.

### III. NEW INSIGHTS AND AREAS OF CONSENSUS

The books we review advance and expand on many of the themes identified in the section above, pushing back on earlier scholarship that argues against the salience of reputation and probing the circumstances under which reputational inferences are most likely to matter. Four of them, Van Jackson's *Rival Reputations: Coercion and Credibility in US-North Korea Relations*, Harvey and Mitton's *Fighting for Credibility: US Reputation and International Politics*, Mark Crescenzi's *Of Friends and Foes: Reputation and Learning in International Politics*, and Danielle Lupton's *Reputation for Resolve: How Leaders Signal Determination in International Politics*, look specifically at reputation, while the fifth, Roseanne McManus's *Statements of Resolve: Achieving Coercive Credibility in International Conflict*, examines credibility of resolve statements. We summarize their contributions below.

For starters, rivalries offer a key opportunity to study whether reputational inferences develop over the course of repeated interactions. Jackson shows that the accrual of reputations between rival states contributes to mutual assessments of threat credibility and affects the likelihood of one of these rivals being challenged by the other in the future, as evidenced in US-North Korea relations over the past fifty years (pp. 15–16). Through several case studies—the *USS Pueblo* crisis in 1968, the EC-121 shoot-down in 1969, the Panmunjom crisis in 1976, and the nuclear crises of the early 1990s and early 2000s—Jackson nicely illustrates a key tension in the effects of acquiring a reputation for resolve versus one for honesty in limited deterrence encounters. The United States established a reputation for honesty but not for resolve with North Korea over a series of interactions in which it declined to use military force in response to provocations, indicating the frank limits of its interests on the Korean Peninsula. For its part, North Korea accrued a reputation for bluffing by only occasionally backing up its incendiary war rhetoric with actual provocations (p. 20).

<sup>18</sup> Lamb 2018; Jervis 1998, 266–71.

Jackson's findings on reputation formation dovetail with those of Harvey and Mitton, who offer a forceful critique of scholarship by reputation skeptics, including Daryl Press, Jonathan Mercer, and Ted Hopf, whom they term the "P-M-H school" (p. 4). Harvey and Mitton suggest that P-M-H produced a false consensus on the irrelevance of reputation during international crises. Through case studies of US deterrence encounters in Bosnia-Herzegovina (1992–95), Kosovo (1998–99), and Iraq (1991–2003), the authors show that actors do indeed look to other states' past actions when making inferences about likely future behavior. Harvey and Mitton take P-M-H to task for use of outdated examples in their case studies, misspecification of rational deterrence theory in developing their arguments, and insufficient probing of discrepant evidence or other mechanisms, aside from past actions, through which reputation may matter. Harvey and Mitton further suggest that the 2013 Syria crisis should be regarded as a coercive success rather than as a failure because the Obama administration ultimately succeeded in brokering a deal with Russia to reduce Bashar Al-Assad's stockpile of chemical weapons.

If past actions are important in structuring reputational inferences, so too are the signals that states and leaders attempt to send regarding their own level of resolve. McManus finds that statements of resolve are most effective when leaders have a clear ability to follow through on them, as demonstrated by American behavior in several Cold War crises (pp. 11–13). In particular, when leaders do not face major obstacles or unacceptable risks in the form of domestic punishment or resistance from veto players, they are more effectively able to convey resolute intentions. Military strength, hawkish domestic veto players, and security in office all contribute to leaders' ability to project resolve (as measured through McManus's novel use of content analysis), but the absence of any one of these factors can undermine statement effectiveness. Case studies of the Cuban Missile Crisis, the refreezing of the Cold War, and the Vietnam War validate her argument; US presidents who enjoyed support from hawkish veto players and the public (John Kennedy and Ronald Reagan) could wield statements of resolve to great effect, while those who lacked these ingredients (Lyndon Johnson and Richard Nixon) could not.

Nevertheless, there is good reason to suspect that reputation is not purely dyadic; signals sent by state A toward state B can also affect judgments made by state C. Crescenzi posits that reputation conditions the environment in which decisions are made because protagonists tend to draw inferences from potential antagonists' behavior toward proxies who are similar to the former in terms of their foreign policy positions,

as measured by a combination of alliances, trade, UN votes, and diplomatic missions, and relative power via the Composite Index of National Capabilities (p. 103). Following the simple “the friend of my friend is my friend” logic, Crescenzi models how extradyadic information helps states draw inferences about others where private information otherwise inhibits credible commitments; a reputation for crisis incompetence yields more violent extradyadic interactions. Crescenzi not only develops these insights via a simple formal model, but also tests his claims through a mix of large-N analysis and historical examples drawn from British alliance formation in the late nineteenth century and following World War II.

While many of the foregoing insights about reputation and credibility portray states as the main actors, Lupton demonstrates that leaders acquire individual reputations for resolve through statements and behavior early in their tenure, but context, including the state’s prior reputation for resolve and its communicated level of interest in the issue under dispute, plays a key mediating role (p. 7). Although the idea that leaders matter is hardly new,<sup>19</sup> Lupton shows through a combination of experiments and case studies that reputations adhere to states mainly in the absence of information about how a new leader is likely to behave. The leader’s initial statements and actions after assuming office powerfully ground others’ expectations for the leader’s future behavior. Lupton’s work is especially innovative for combining a microfoundational perspective on her research question, through process tracing survey experiments that manipulate key features of both context and leader behavior, with case studies that probe how Soviet premier Nikita Khrushchev evaluated two US presidents—Dwight Eisenhower and Kennedy—through a reputational lens.

If credibility is some combination of capabilities, interests, and reputation (for resolve), we diagram the authors’ contributions as follows in Table 1 and Figure 1.

The books we review significantly advance our understanding of when and how reputation matters in international relations, even though they focus on different dependent variables, independent variables, and scope conditions, while also taking a variety of approaches regarding to whom reputation adheres. For example, Jackson’s analysis deals squarely with reputation between states and does not attempt to evaluate the role of leaders in the formation of reputation. For Harvey and Mitton, reputation and credibility are also assigned to states, although the extent

<sup>19</sup>Byman and Pollack 2001; Guisinger and Smith 2002; McGillivray and Stam 2004; Wolford 2007; Goemans, Gleditsch, and Chiozza 2009.



TABLE 1  
MIXED-METHODS APPROACHES

	<i>Methods</i>	<i>Unit of Analysis</i>	<i>Dependent Variable</i>	<i>When Reputation or Resolve Matters</i>
<i>Jackson</i>	case studies	dyadic, states	reputational inferences in U.S.–N.K. dyad	in rivalries
<i>Harvey and Mitton</i>	case studies	dyadic, states	reputational inferences about U.S. by others	in degree of cross-case similarity
<i>McManus</i>	case studies, content analysis, game theory, large-N quantitative	dyadic, mix of leaders and states	dispute outcomes	in absence of veto players and domestic constraints
<i>Crescenzi</i>	case studies, game theory, large-N quantitative	dyadic/triadic, mix of leaders and states	reputational inferences toward third parties	in degree of cross-case similarity
<i>Lupton</i>	case studies, experiments	dyadic, leaders	reputational inferences about leaders	for states, early in leader's tenure; for leaders, with time in office

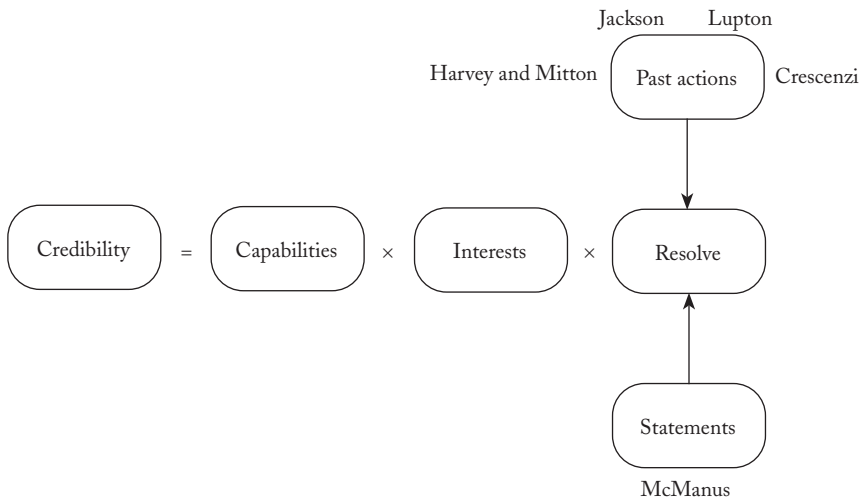


FIGURE 1  
DIAGRAMMING NEW CONTRIBUTIONS

to which they are transferrable depends heavily on the similarities between past and current events. In that long list of potential variables, Harvey and Mitton include but do not attempt to unpack the role of leaders. This contrasts with Lupton's argument that observers rely on state reputations only for the relatively short period when a new leader comes to power, modifying or replacing it as they come to see how the new leader behaves in various situations, at which point reputation becomes more leader specific. For additional contrast, Crescenzi's theoretical model also includes learning, but unlike in Lupton's work, this phenomenon takes place between states rather than leaders.

Although the books under review are diverse in their specific research questions and methodological approaches, they have achieved consensus on two important matters—treating reputation as a relational concept and delineating scope conditions on when reputational inferences are likely to matter.

#### IN THE EYE OF THE BEHOLDER

Naïve observers often compare international politics to chess or poker, but doing so is misguided; not only are the rules of both games set ahead of time but also all the moves are out in the open and participants start from a common understanding of how to play. A more accurate comparison is to the Akira Kurosawa film, *Rashomon*, in which each participant sees a situation and what everyone is doing quite differently and fails to realize that his or her understanding is not shared by others. In reality, signaling and perception are two sides of the same coin, but signals are only meaningful if the observer understands what the signaler is trying to say (or if the observer even realizes that a signal is being sent in the first place). As a deception planner in World War II explained to his subordinates when selecting the material that would accompany a corpse to be floated to the shore in Spain, where it was certain to be turned over to the Germans, "What you, a Briton with a British background, think can be deduced from a document does not matter. It is what your opposite number, with his German knowledge and background, will think that matters. . . ." <sup>20</sup>

Contrary to what many decision makers may believe about reputation and credibility, the two are relational qualities that actors do not own for themselves. The observer's judgments about the actor are those that matter for the latter's reputation, and try as the actor might, he or she has only limited ability to influence what the observer is think-

<sup>20</sup>Jervis 1970, 70.

ing.<sup>21</sup> States aim to create desired impressions on others, but this goal often fails because of the others' mindsets. The challenge in creating these desired impressions means that scholarly theories about how signals should be perceived can run aground on the way the observer interprets them.<sup>22</sup> For example, one of the main reasons why the United States fought in Vietnam was to bolster its reputation for protecting allies. But most allies did not see it that way and neither supported the war nor thought that America's being willing to fight it indicated that it would come to their aid.<sup>23</sup> To truly understand why communication between states succeeds or fails, we need theory and empirics that examine what the signaler is trying to say, how the perceiver interprets it, and over what duration these impressions last.

Recent studies embrace this perspective to good effect, allowing us to learn more about the scope conditions under which reputations form and change (discussed in more detail below). As Jackson notes, "Because credibility and reputation are relational concepts, they depend entirely on the perception of others" (p. 20). Harvey and Mitton agree that "credibility is both complex (i.e. multifaceted) and in the eye of the beholder" (p. 28). Crescenzi also points out that reputation is "contextual and in the eyes of the beholder. Because of this, reputational information can influence one actor's behavior while being completely dismissed by another" (p. 28). Lupton nicely summarizes the key insight: "As reputations are rooted in one actor's perceptions of another, we cannot simply look at a leader's behavior and simply assume that such behavior will be interpreted in a certain way by others" (p. 23). But limitations in the documentary record can make it difficult to gather the most relevant details about the observer's point of view. Therefore, although McManus argues that security in office and a lack of veto points enhance the credibility of resolve statements from US leaders (pp. 28–34), she is able to do a much better job of measuring these variables directly than of showing how adversaries perceived them, which is crucial for her argument.

In addition, by taking the observer's perspective seriously, the books under review are able to offer a more nuanced interpretation of how reputational beliefs form in response to actions versus words. Some of these insights are quite novel and have important policy implications. For example, Jackson finds through his case studies that US officials

<sup>21</sup> Schelling is in many ways progenitor of the work on credibility and reputation, but he is very aware of this: see, for example, Schelling 1966, 56–57.

<sup>22</sup> Jervis 1976; Jervis 2017, chap. 5.

<sup>23</sup> Logevall 1999.

routinely discounted overheated rhetoric from North Korea in the absence of follow-through and only took North Korean threats seriously after a provocation, such as the seizure of the surveillance ship *Pueblo*, while North Korea similarly seemed to have interpreted US inaction in response to provocations, such as its failure to retaliate for the ship's seizure as green-lighting subsequent escalation, including the shoot-down of the EC-121 (p. 137). Strong actions by North Korea then induced a sudden and drastic change in American perceptions. This case raises historical parallels with the Cuban Missile Crisis, in which Khrushchev discounted Kennedy's statements that he would not permit the Soviet Union to station offensive missiles in Cuba and only drew back when the United States took strong and risky actions after the missiles were discovered.

Three of the titles also integrate evidence from the observer's perspective to show that there is a critical interaction between statements and behavior in generating reputational inferences. Crescenzi shows that potential antagonists' behavior toward proxies—particularly the antagonists' reputation for crisis incompetence (by which he means trustworthiness) vis-à-vis similar states—is the key element that observers tend to watch (pp. 23, 88–91). In debating whether to ally with Germany or Japan at the turn of the twentieth century, British decision makers highlighted Germany's recent history of unreliability and poor treatment of alliance partners with regard to preventing Russian incursions in China, against Japan's sterling record in quelling the Boxer Rebellion and defeating China in the 1902 Sino-Japanese War. Harvey and Mitton argue that for a favorable outcome, statements of resolve must be backed up by threats and assurances that are credible or that involve clearly communicated statements of interest and the capability to follow through. When the United States failed to respond to probes and challenges from the Bosnian Serbs, violence escalated (p. 108). Saddam Hussein gathered from repeated cycles of violence short of war that the United States would not wage a ground campaign against Iraq for fear of American casualties. McManus echoes this finding. Her combination of content and statistical analysis shows that highly resolute statements increase the odds of favorable dispute outcomes, especially when coupled with factors that suggest the capacity and willingness to act; these include a preponderance of military capabilities, security in office, and support from hawks in Congress (p. 92).

While the accounts above rely on case studies to uncover the observer's perspective, Lupton complements her qualitative analysis with experimental evidence on the interaction between statements and behavior in shaping reputational dynamics. Lupton finds that leaders'

statements and behavior intermingle with states' prior actions and their communication of interest in the matter at hand to shape observers' perceptions of resolve (p. 70). Because observational data rarely reveal the relative impact of the state's behavior compared to its statements, survey experiments can be especially helpful. Indeed, Lupton is able to randomize variation in context and leader-specific information to isolate the effects of the former (past state behavior, military capabilities, and level of interest) from the latter (leader statements and past behavior) on observers' estimates of leaders' toughness and determination. Her results suggest that leaders' irresolute statements and behavior have negative effects of roughly equal magnitude on perceived resolve, even after controlling for the various contextual factors.

Last, by focusing on observers' assessments, these scholars find that judgments about credibility and reputation tend to be quite sticky. For Harvey and Mitton, this is true of the United States' limited deterrence encounters in the Balkans, Iraq, and Syria since the end of the Cold War because the type of opposing regimes and the nature of the conflicts in question approximate one another (pp. 88–89). Crescenzi concurs, but models a form of exponential decay in these effects over time: past actions matter and their influence becomes less (more) informative as time passes without (with) new instances of cooperation or conflict (p. 50). Lupton, too, finds that these assessments are quite difficult to alter once established, as statements and behavior that come early in a leader's tenure figure prominently in observers' expectations about future interactions, while background or contextual information tends to decline in relevance (p. 77). All of this suggests that states must go to great lengths to convince others to adopt a favorable view of their own behavior. But as Crescenzi further shows, garnering a reputation for hostility, violence, or escalation may not be such a good thing and can even backfire, at least to the extent that such qualities may convince opponents that they themselves need to escalate (pp. 88–89). This evidence provides important confirmation of what was already known in part—that states often want to be perceived as being tough in the sense of not readily backing down but must balance this desire against becoming known for unreasonable, aggressive, or bullying actions.<sup>24</sup>

### SCOPE CONDITIONS

A second accomplishment in recent work on reputation has been to move past earlier debates on whether reputation matters at all and to focus instead on the relevant scope conditions for when reputation is likely to be a considered factor. This welcome development follows a

<sup>24</sup> Snyder and Diesing 1977.

long stalemate between several older studies that cast doubt on reputation formation in international politics<sup>25</sup> and subsequent work that has pushed back on that assertion.<sup>26</sup> Most of the books under review find that reputation does matter, and show that decision makers in state A do take account of state B's past actions when assessing the latter's intentions; these scholars take their conclusions a step further by exploring the situations under which these inferences are most salient: under rivalry (Jackson), in similar types of conflicts with similar types of states (Harvey and Mitton, and Crescenzi), and in the absence of leader-specific information at the beginning of that time in office (Lupton).

First, as Jackson posits, rivalry is ripe for studying reputation precisely because it features conditions of endemic hostility over repeated interactions. For those who might dispute whether US-North Korea relations truly constitute a rivalry, Jackson posits that rivalries do not require power parity but do involve several other key parameters that are likely to exacerbate reputational dynamics, including competing claims over legitimacy and acceptable behavior as international actors; the attachment of heightened meaning to otherwise low-salience issues and ambiguous statements or actions, which can trigger escalation by reinforcing existing biases; and persistent anticipation of future disputes (pp. 7–8). Although these are not wholly new points,<sup>27</sup> they accurately characterize the contours of US-North Korea relations—the parties disagree over the legitimacy of South Korea's existence as well as about the continued exclusion of North Korea from most international institutions. Seemingly minor incidents, such as the tree-chopping episode at Panmunjom in 1976, have nearly escalated into war while expectations of eventual conflict have remained high throughout recurring crises. As a result, this dyad provides fertile ground for testing common deterrence-based arguments about the role of reputation.

Second, reputational inferences are likely to be influential when similarities between cases are evident. As Harvey and Mitton suggest, past actions are neither definitive for reputation (because cases often differ in important ways) nor completely irrelevant (given that lessons learned from previous encounters frequently enter into subsequent calculations about credibility). Their argument indicates that leaders like Ratko Mladic, Slobodan Milosevic, and Saddam Hussein were willing to probe US and allied resolve at least in part based on evidence from within and across the Bosnia, Kosovo, and Iraq cases (pp.105–107). In

<sup>25</sup> Hopf 1994; Mercer 1996; Press 2005.

<sup>26</sup> Weisiger and Yarhi-Milo 2015.

<sup>27</sup> Huth 1997.

Kosovo, for example, NATO's initial show of force during the spring and summer of 1998—through air exercises over Albania and Macedonia and subsequent verbal threats from the United States about intervening—were insufficient to deter Serbian security forces from continuing their crackdown on ethnic Albanians. It was not until roughly a year later, after NATO's seventy-eight-day bombing campaign, that Milosevic capitulated. Harvey and Mitton argue that it took Milosevic so long to give in not just because of NATO's failure to back up threats with action, but also because of the general reputation that the United States had apparently garnered by resorting only to limited air strikes against Iraq in Operation Desert Fox during December 1998 (pp. 129–31). This evidence may have led Milosevic to conclude that any bombing campaign would be short and that he could undermine American resolve through a willingness to take casualties among his own forces.

Crescenzi formalizes this insight. When private information makes it difficult to judge whether others' commitments are credible, protagonist states rely on historical information from beyond the dyad, that is, protagonists look at how antagonists treat states of similar foreign policy orientation and relative power (pp. 40–43). In Crescenzi's work, reputational learning was key in shaping the United Kingdom's foreign policy toward the Soviet Union after World War II via British trust in and reliance on American expertise: in Moscow during early 1946, British Chargé d'Affairs Frank Roberts leaned on George Kennan's famous assessment of Soviet intentions in advising London on how to handle the USSR (p. 52).

Third, although reputation is fundamentally a judgment about disposition, situational factors can still affect its relative salience. On the one hand, Lupton finds that over a leader's tenure in office, the prominence of context-dependent, state-level qualities (capabilities, interests, and past behavior) declines while the relevance of leader-specific attributes (statements and behavior) rises. Looking at two American presidents through Khrushchev's eyes, Lupton shows that Eisenhower succeeded where Kennedy failed in establishing a reputation for resolve during successive crises over Berlin. Whereas Ike conveyed to Khrushchev that he would not make major concessions on the issue and backed up this position with repeated firm statements, JFK's twin disasters at the Bay of Pigs and the 1961 Vienna summit suggested to Khrushchev that he would not follow up his strong rhetoric with firm action. Kennedy therefore did not benefit much from whatever reputation Eisenhower had established for himself or the United States because Kennedy's initial statements and behavior led Khrushchev to

conclude that he was irresolute. On the other hand, Jackson demonstrates that one reason why the North Koreans felt free to engage very provocatively in the late 1960s, most obviously in seizing the *Pueblo* and shooting down the American surveillance plane the next year, was not that they judged the United States to be generally lacking in resolve, but that they correctly thought the US would be unwilling to retaliate while it was so deeply mired in Vietnam. In this case, contextual factors may have superseded judgments about type.

#### IV. DIRECTIONS FOR FUTURE RESEARCH

Despite substantial progress in this research agenda, there are several areas that would benefit from increased focus in future research, including greater attention to methodological trade-offs and analytical clarity about the unit of analysis, additional theorizing on differences among observers, the role of psychology in shaping reputational beliefs and credibility assessments, and the need for further unpacking of the multidimensionality of reputations and the trade-off between different types of reputations.

##### METHODOLOGICAL AND ANALYTICAL CLARITY

Scholars across the discipline increasingly employ a large toolkit in service of their inquiries. The rise of mixed-methods approaches has had important implications for the study of credibility and reputation. We not only observe support for reputational inferences coming from both qualitative and quantitative research, but also see how scholars have leveraged the unique advantages of particular methods to circumvent some of the limitations inherent in purely qualitative studies. The rich mixture of empirical approaches on display in the books we review, encompassing case studies, experiments, game theory, and large-N quantitative analyses, expands the possibilities for what we can learn by paying closer attention to leaders' public statements, by deploying experimental techniques to more neatly identify causal mechanisms, and by using relatively simple models to derive testable intuitions and predictions.

Nevertheless, the books vary in the extent to which they offer empirical evidence in support of their hypotheses and some of the authors are better than the others in acknowledging the limitations inherent in their methods of choice. A core methodological challenge in studying credibility and reputation lies in choosing whether to focus on how reputations change or remain steady over time as each state learns or



fails to learn from succeeding interactions, on the one hand, or to treat cases as independent from each other and to determine the variables that distinguish outcomes, on the other. Regardless of their specific method, most studies can only examine snapshots in time. For example, survey experiments that draw on non-elite samples are a welcome tool for establishing behavioral baselines. But even though recent work shows broad similarities in how elites and publics think about foreign policy questions,<sup>28</sup> the measurements are taken at a point in time and under certain conditions that may not always generalize. Moreover, although survey experiments can analyze the microfoundations underlying many theories of credibility and reputation, they do not always capture the information-rich environment that real leaders face. For example, Lupton's surveys manipulate past actions and only one other contextual factor at a time. As a result, her experiments may oversimplify what is really a complex strategic interaction with multiple relevant variables.<sup>29</sup>

A related issue concerns case selection, particularly with respect to what examples are representative of how states and leaders assess reputation and what level of temporal distance is appropriate for inclusion in a study. Our understanding of history is further skewed because we know more about cases in which credibility is underestimated than about those in which it is overestimated because the former often lead to dramatic conflict and the latter lead to nonevents or perhaps to retreats that do not immediately produce vivid outcomes. Similarly, the case study approach is subject to a number of limitations, including the familiar trade-off between depth and breadth. Even when using archival documents, potential pitfalls remain. While the historical record is rarely beyond debate, critiques of earlier work that finds no evidence for discussion of reputation among decision makers note that reliance on historical documents to study reputational dynamics may lead scholars to overlook implicit, unspoken assumptions, especially if decision makers view reputation as so evidently important as to not be worth mentioning.<sup>30</sup> An additional concern, especially in cases involving the United States, arises from the imbalance of available evidence on what American decision makers were trying to achieve and how they judged others (where primary and secondary sources are abundant) versus oth-

<sup>28</sup> Yarhi-Milo, Kertzer, and Renshon 2018; Kertzer, Renshon, and Yarhi-Milo 2019; Kertzer Forthcoming.

<sup>29</sup> See Kertzer, Renshon, and Yarhi-Milo 2019 for an information-rich conjoint experiment in which past action is manipulated alongside several other country-level and leader-level variables.

<sup>30</sup> See Weisiger and Yarhi-Milo 2015, 8–9, for more on this critique. Also see Hopf 1994; Mercer 1996; Press 2005; Dafoe, Renshon, and Huth 2014. The most famous discussion is Joll 1968.

ers' perceptions and judgments of American signals and actions (where primary sources are few and secondary sources are neither directly focused on these questions nor necessarily drawing on the best possible evidence). A true grasp of how American credibility and reputation form and change remains elusive in the absence of an improved documentary record.

Furthermore, if the use of archival documents requires the selection of historically bounded episodes (crises or conflicts) that are far enough in the past to make process tracing tractable, qualitative approaches may necessarily select on exceptional circumstances or generate scope conditions with limited external validity.<sup>31</sup> For instance, Harvey and Mitton suggest that the evidence marshaled by Press, Mercer, and Hopf relies on outdated examples from the early twentieth century, the world wars, and the Vietnam War that may not be applicable to more temporally proximate, limited deterrence encounters. Other issues emerge when we consider that actors behave strategically: they may attempt to cast their decisions in a favorable light by offering justifications for policy choices that would look sensible if revealed publicly but may not reflect their actual beliefs. For example, the "Munich analogy" carried substantial weight among US government officials who were determined to avoid a reputation for appeasing aggression during the Cold War.<sup>32</sup> But when scholars look back at invocations of this analogy, they must keep in mind actors' strategic incentives because it may be difficult or impossible to tell whether policymakers marshaled arguments about credibility and reputation sincerely, used them as a cudgel to get their preferred solutions adopted, or found themselves in the throes of groupthink.<sup>33</sup>

The methodological challenges that scholars face in studying reputation and credibility are intimately linked to the unit of analysis at which their theory is operating. Although most scholars still implicitly or explicitly assume that reputations and credibility adhere to states, there is a growing recognition that both operate (differently) at different levels. However theoretically and empirically valid, this development presents more questions than answers regarding how to synthesize insights from different levels of analysis or how new findings interact with previous research on the determinants of leader behavior.<sup>34</sup> Although work by Jonathan Renshon and colleagues as well as by Cathy Wu and Scott Wolford has laudably sought to integrate leader- and state-specific

<sup>31</sup> Marinov 2005; Dafoe, Renshon, and Huth 2014, 184–85.

<sup>32</sup> Khong 1992.

<sup>33</sup> Janis 1972.

<sup>34</sup> Byman and Pollack 2001; Wolford 2007; Jervis 2013; Horowitz, Stam, and Ellis 2015.

variables in the study of reputation,<sup>35</sup> the research program still lacks an overarching framework for theory development. For instance, how should we treat the reputations of substate actors like particular bureaucracies or individual advisers? And if different reputations are assigned to different actors, do observers attempt to aggregate them into a coherent whole, or do they focus on particular individuals or organizations?

The answers to these questions matter because the extent to which reputation or credibility adheres to states, leaders, or other entities carries significant implications for both theoretical development and research design. Theoretically, models of reputation formation among leaders emphasize different types of indicators as compared to state-specific factors. For instance, leader-level theories stress characteristics like combat experience, tenure in office, personality traits, and gender while highlighting idiosyncratic factors that historians love and that frustrate political scientists' attempts to generalize, whereas state-level theories focus on factors like capabilities, interests, and regime type.<sup>36</sup>

To highlight the importance of the unit of analysis, consider the following contemporary example. If John Bolton and other observers of President Donald Trump are right, Trump is unique among American leaders in believing that foreign dictators are particularly prone to keeping their word. Therefore, the stability of the perceiver's beliefs about reputation over time likely hinges on whether the perceiver believes that reputation and credibility assign to states or to leaders. If reputation adheres to states, then perceivers ought to hold quite sticky beliefs in the absence of some radical discontinuity—as Jackson finds in the rivalry between North Korea and the United States. But if reputation applies to leaders, it should exhibit substantially more variation and potentially shift every time a new leader comes to power, as Lupton explores. Renshon and colleagues' findings indicate that state-specific effects are about double the size of leader-specific effects.<sup>37</sup> A key policy implication, then, is that leaders may either be able to hide behind or feel forced to run from their state's existing reputation, but it will likely take significant effort to convince perceivers to update their beliefs accordingly.

#### DIFFERENCES AMONG OBSERVERS

Although we know that individual perceptions color beliefs, the books reviewed are not designed to unpack how homogeneous or heteroge-

<sup>35</sup> Renshon, Dafoe, and Huth 2018; Wu and Wolford 2018.

<sup>36</sup> Kertzer, Renshon, and Yarhi-Milo 2019, 4.

<sup>37</sup> Renshon, Dafoe, and Huth 2018.

neous these perceptions are within and across states. The studies stop short of observing or explaining whether such variation in concern for credibility at the level of leaders, bureaucracies, or regime types even exists, and if it does, what it looks like and what factors are driving such dynamics. We know that beliefs held by leaders, organizations, or particular types of regimes shape information intake, receptivity to alternative explanations, and ultimate inferences, so there is good reason to suspect that they also play a powerful role in shaping observers' reputational inferences and assessments of credibility, even in the face of situational constraints.<sup>38</sup> This may be true for several reasons.

First, the tendency for observers to assimilate new information into preexisting beliefs may play an important role in sustaining reputation. One reason why President Barack Obama's failure to use force when Assad crossed the "red line" and used chemical weapons earned him a reputation for being weak, despite the fact that he and Russia were able to broker an agreement that removed most of Syria's chemical stockpile, was that many audiences were predisposed to see him as indecisive. By contrast, Trump's two strikes in response to Assad's use of chemicals combined with his belligerent style meant that he was not uniformly seen as weak when he failed to act in the face of credible reports that the Syrian president was continuing to use poison gas, and later called off a retaliatory bombing attack against Iran shot down an American drone in June 2019. What matters is the observer's preexisting beliefs about what she is seeing. Future research may examine when and how quickly updating from different reputational baselines occurs, whether for resolve or some other quality, and how preexisting beliefs interact with new information in this process.

Second, idiosyncratic differences among observers can significantly shape assessments of reputation and credibility in ways that actors may fail to appreciate or anticipate. For example, it appears that the Soviets were very impressed by President Ronald Reagan's resolve because he fired the air traffic controllers when they went on strike in 1981,<sup>39</sup> while most other observers, whether they approved of Reagan's actions or not, did not see this as indicating characteristics that would display themselves in latter foreign policies. Almost twenty years earlier most observers were impressed by Kennedy's resolve (and diplomatic ingenuity) in the Cuban Missile Crisis, but what greatly bolstered his repu-

<sup>38</sup>These examples remind us of the difficulties in applying standard Bayesian updating when the prior beliefs strongly color the perceived diagnosticity of the new information. For a critique of attempts to use standard Bayesian updating techniques in modeling how prior beliefs affect inferences, see Jervis 2017, xlvii-lii.

<sup>39</sup>Shultz 1996, 1135.

tation with Khrushchev was his imputed ability to resist the belligerent demands of the military. This told the Soviet leader that Kennedy not only sought better relations with the USSR, but also would be able to prevail in his struggle with the warmongers—reminding us that actors make subjective inferences influenced by their theories about the how the world works, what the other is like, and what seems most vivid to them.<sup>40</sup> Most Americans knew that Kennedy was in charge and so were not impressed in this way; Khrushchev's views of Kennedy and the US political system were different, and so he was.

A more current example further illustrates this point. When the negotiations between the United States and South Korea about how much the latter would pay for US defense broke down in mid-November 2019, would North Korea (and others) infer that the United States was less likely to come to the South's assistance in the event of invasion because the breakdown showed both significant conflict between the United States and the South and the limits on the importance that the US placed on that country? Or would observers think that it showed how tough Trump was? Logic will not provide the answers. We need to know, or at least to estimate, Kim Jong Un's understanding of the world and the United States.

Third, because reputation is a belief, states may attempt to influence or manipulate their reputation with the explanation they give for their actions. Because behavior does not speak for itself, reputation depends in part on why receivers believe the state has acted as it did. Did it back down because of factors that are only temporary? Was it seeking not to avoid war at all costs, but to obey international law or some other general principle or, more pragmatically, to set relations with the adversary on a better footing? Taking advantage of the fact that observers have to do important interpretive work in divining the other side's expectations, theories of communication, and political or personal motives, the sender can sometimes attempt to manipulate observers' reputational inferences by offering a coherent explanation for its behavior. When others' images are not firmly established, observers may search for signposts and frameworks that can guide understanding of what the other is doing. But states, like individuals, sometimes provide excuses for undesired behavior. To avoid a reputation for being unreliable, for example, a state may point to factors that made it impossible for it to do as it said it would but that will not recur in the future and so should not affect its reputation. Actors who escalate but who do not want to be seen as rash could explain that the situation was uniquely provok-

<sup>40</sup>Yarhi-Milo 2014.

ing, that any leader would have behaved in the same way, and that no broader inferences are called for.

How, then, might observers react to behavior that actors label as important for their reputations? Jon Elster writes that it is obvious that "Nothing is so unimpressive as behavior designed to impress."<sup>41</sup> But this observation is not entirely true in foreign policy. Even if the manipulating actor can never fully control what conclusions others will draw, displays that are calculated may be believed to create more of a commitment and to reveal more about how an actor will behave in the future than is the case for spontaneous reactions, and in some cases states explicitly say that they are trying to impress.<sup>42</sup> This behavior is a common habit in American foreign policy, with leaders frequently maintaining that some costly and often unpopular policy is needed to maintain American resolve and to bolster its credibility; it was the main rationale for fighting in Vietnam and resisting Soviet pressure to withdraw from Berlin during the Cold War, to take only two of the most obvious examples. And still, scholars have yet to examine whether observers' inferences are affected by the actor's stated rationale as well as her behavior. Do statements and actions reinforce each other? Or do claims that the state is behaving in a particular manner to provide a desired impression undercut that very impression?

In sum, the discussion above highlights why studying reputation and credibility requires scholars to look at multiple observers within or across states. Only when we have evidence that observers draw the same reputational inferences from the actor's behavior can we neglect factors that differentiate one observer from another. Yet because much of the literature on these topics concentrates on the inferences of one actor only, we have not been able to fully wrestle with the question of how generalizable our theories are. For instance, while Lupton nicely unpacks Khrushchev's impressions of Eisenhower and Kennedy, the other side of this coin is that each president assigned the Soviet leader a different reputation. By 1960, Eisenhower had become convinced that Khrushchev's lack of actions to fulfill his threats on Berlin meant that he was bluffing and that neither concessions nor military preparations were necessary; Kennedy was less certain that Khrushchev would continue to acquiesce in the status quo. Along similar lines, one of the signal contributions of Mercer's study is that he shows how allies and adversaries draw different inferences from the behavior of a target state. Likewise, while Crescenzi indicates that an antagonist's behavior to-

<sup>41</sup> Elster 1983, 66.

<sup>42</sup> Goffman 1959; Jervis 1970.

ward a protagonist's proxy will influence the protagonist's views of the antagonist (p. 45), it is unclear how widely shared these assessments are within each observing state.

### THE PSYCHOLOGY OF THE PERCEIVER

One possible lens that scholarship on reputation and credibility can use to systematically evaluate differences among individuals or groups of actors is to more explicitly and systematically incorporate insights from psychology. Such insights can offer an opportunity to either test these explanations against rationalist approaches or to synthesize theory that integrates both perspectives.<sup>43</sup> Indeed, research involving social psychology, behavioral economics, and neuroscience is currently experiencing a renaissance in the study of international politics.<sup>44</sup> Some of the insights generated in those fields can be more meaningfully integrated into the study of reputation and credibility. Here, we offer ways scholars can build on those emerging insights.

Emotions, in particular, are foundational to beliefs and may explain not only why actors worry about reputation in the first place, but also how they process costly signals, as Mercer has shown regarding the behavior of President Harry Truman and Secretary of State Dean Acheson at the outset of the Korean War.<sup>45</sup> Indeed, emotions are integral to how our brains update in response to new information. Specifically, there is evidence to suggest that we are more likely to update our beliefs in response to positive rather than to negative news about the world as well as about ourselves.<sup>46</sup> This predilection may have important implications for how leaders process incoming positive versus negative information about their own reputation, such as input from intelligence advisers on how other states or leaders view them, or the likely behavior of others.

More broadly, there remains somewhat of a disconnect between advances in the literature on leaders and decision-making and the scholarship on reputation and credibility. Fundamentally, the reputational beliefs that leaders hold are influenced, like other beliefs, by very human personal, psychological, and political needs. Yet current studies have not sufficiently explored how needs for cognitive closure or avoidance of psychological stress shape observers' reading of others' reputations. For example, it is likely that one reason why Khrushchev saw

<sup>43</sup> For an example of rationalist-psychological fusion, see Mintz 2004.

<sup>44</sup> Holmes 2018; Markwica 2018; Wheeler 2018; Davis and McDermott 2020.

<sup>45</sup> Mercer 2010; Mercer 2013a.

<sup>46</sup> Sharot and Garrett 2016; Kuzmanovic and Rigoux 2017.

Kennedy as weak in the period leading up to the Cuban Missile Crisis is that he was in a very difficult situation. Not only was his intercontinental ballistic missile force lagging, but he also knew that Kennedy (and the entire world) knew this. Putting missiles into Cuba was a very attractive solution, and doing so led Khrushchev to see Kennedy as the kind of leader who would let him get away with it. President Franklin Roosevelt's belief that Stalin was a leader who, although difficult and suspicious, could be brought into a peaceful world order similarly owed something to his realization that the United States would face a dire situation if Stalin were a different type.

Moreover, while a host of recent research has pointed to leader-specific variables, such as combat experience,<sup>47</sup> age,<sup>48</sup> gender,<sup>49</sup> and personality traits, such as narcissism, that affect leader behavior in militarized disputes, the link between these studies and research on credibility and reputation remains largely elusive.<sup>50</sup> Scholarship could benefit from examining psychological constructs that can systematically shape beliefs about reputation and credibility. For example, if it is true, as James Davis and Rose McDermott suggest, that "as individuals age, changes in the influence of emotions, motivation, and energy level on decision-making appear to decrease their sensitivity to anticipated loss,"<sup>51</sup> then scholars could examine whether leaders' concern for losing reputation (for resolve, for instance) decreases as they grow older.

Leaders' predispositions can also be studied more systematically to account for previously puzzling differences in reputational beliefs among leaders, elites, or citizens. Recent work has begun to integrate those perspectives. For example, on the issue of domestic reputational costs, Ryan Brutger and Joshua Kertzer find that citizens' predispositions about the utility of military force shape concerns for different types of reputations; hawks (doves) seem to be concerned about the negative reputational consequences of inconsistency (belligerence or interventionism).<sup>52</sup> Leaders' personality traits, specifically their inclination toward self-monitoring, systematically explain variation in leaders' concerns about reputation for resolve as well as their willingness to fight for face.<sup>53</sup> On the study of credibility, survey experiments by Keren Yarhi-Milo, Kertzer, and Renshon have shown that decision makers

<sup>47</sup> Horowitz, Stam, and Ellis 2015.

<sup>48</sup> Davis and McDermott 2020.

<sup>49</sup> Schramm and Stark 2020.

<sup>50</sup> Although see Schwartz and Blair 2020.

<sup>51</sup> Davis and McDermott 2020; Samanez-Larkin and Knutson 2015.

<sup>52</sup> Brutger and Kertzer 2018.

<sup>53</sup> Yarhi-Milo 2018.



and individuals who display hawkish or dovish tendencies on foreign policy issues exhibit different interpretations of the credibility of costly signals like military mobilization and public threats of escalation.<sup>54</sup> Indeed, costly signals of reassurance are only interpreted as such by those whose predispositions already incline them toward cooperative internationalism—in other words, by individuals who are least in need of reassurance in the first place.<sup>55</sup>

But overall, the vast majority of the literature, including the books we consider here, do not integrate these unit-level psychological dynamics in their examination of reputation or credibility, perhaps in part because for many of the authors, the perceiver they study is a state. Even models that conceive of reputation formation between leaders (Lupton) or of resolve as a function of leaders' statements (McManus) do not examine how predispositions of the leaders themselves may shape reputational learning or credibility inferences. In these accounts, respectively, Khrushchev apparently arrived at his estimates of Eisenhower as strong and Kennedy as weak according to the substance of his interactions with each and with dispositional factors playing no particular role, while the Soviet and North Vietnamese leadership reacted essentially uniformly to American statements of resolve. In contrast, Hopf finds evidence of contestation within the Politburo over what the Vietnam War meant for American credibility.<sup>56</sup> Bringing psychological variables to bear could help to explain these types of discrepancies and to better connect the study of reputation and credibility with recent advances in behavioral approaches to international relations.<sup>57</sup>

Last, psychological perspectives can help to address tensions over when observers draw situational (that is, features of the context or environment) versus dispositional (that is, a leader or state's type) inferences about the credibility and reputation of others.<sup>58</sup> At one extreme, we can imagine a world in which situational variables are never important and dispositional factors are fully determinative. This scenario would imply a world of strong interdependence in which type dictates behavior and states price this expectation into their interactions. At the other extreme, we can envision a world in which situational factors always trump dispositional ones and events are independent. In

<sup>54</sup> Yarhi-Milo, Kertzer, and Renshon 2018; Kertzer, Renshon, and Yarhi-Milo 2019.

<sup>55</sup> Kertzer, Rathbun, and Rathbun 2020.

<sup>56</sup> Hopf 1994.

<sup>57</sup> Kertzer 2016; Hafner-Burton et al. 2017.

<sup>58</sup> We distinguish here between dispositional attributions, or judgments about another's type, and predispositions, which are a feature of actors' psychologies that may make them care more or less about reputation, as in the discussion above about costly signals.

such a case, states would not transfer information about the other actor's type from one interaction to the next. The books we review offer some evidence that states or leaders will draw sticky dispositional inferences on the basis of past behavior when assessing another's credibility, especially over repeated interactions and in similar contexts. But psychological insights might allow us to better identify when dispositional versus situational attributions are more likely and therefore, the conditions under which reputations are likely to persist or to evolve. For instance, deriving a measure of individual sensitivity to the interdependence of events could help to establish what types of leaders are more likely to characterize others' behavior in situational versus dispositional terms. The more that observers see events as interconnected, the more we might expect them to use type rather than context to explain others' behavior.<sup>59</sup>

But the task of parsing situational and dispositional factors is complicated by what psychologists call the "fundamental attribution error," or the tendency of observers to overweight the role of dispositions and underweight the role of the situation when explaining how others behave. In international, and perhaps interpersonal, behavior, an exception seems to be that when the other behaves in a way that conforms to the observer's attempted influence or desired behavior, situational factors in the form of what the observer did are given great weight.<sup>60</sup> As scholars of credibility and reputation, we must be sensitive to these contingencies and avoid explanations that portray observers as only making one type of inference (whether dispositional or situational). The books reviewed here mostly concern observers' judgments about type (Crescenzi, Jackson, and Harvey and Mitton), although some are attentive to both how observers categorize others' types and what contextual factors, such as military capabilities (Lupton) or domestic political constraints (McManus), are also relevant. The future of this research program almost certainly lies in theory and empirics that grapple with and find ways to reconcile the relative influence of situational and dispositional factors among observers.

#### MULTIDIMENSIONALITY AND TRADE-OFFS

Most scholarship in international politics still refers to reputation and credibility as single entities, but it would be better to talk about credibility and reputation for what, as well as about the trade-offs that decision makers may face in pursuing different types of reputations before

<sup>59</sup>Jervis and Snyder 1991.

<sup>60</sup>Jervis 1970, chap. 9; Mercer 1996.

different audiences.<sup>61</sup> Scholarship in international security has focused on credibility and reputation in the context of threats, though notable exceptions do exist. For example, work by Anne Sartori and Crescenzi examines reputation for attributes like honesty and upholding commitments.<sup>62</sup> Other scholars, including Gregory Miller and Douglas Gibler, zoom in on reputation for alliance reliability while the wider literature suggests that states and leaders can acquire reputations for qualities like integrity, competence, safeguarding honor, impetuosity, violence, hostility, dealing with secessionists, and imposing economic sanctions.<sup>63</sup> But even though the difficulty and importance of making promises credible is also central to our understanding of reputation, these topics have received far less attention.<sup>64</sup> What, if any, are the linkages between a reputation for living up to one's threats and reputation for fulfilling one's promises? Are these independent? Or do they tap beliefs about general characteristics of the actor's signaling or reputation?<sup>65</sup> The need to make promises credible is understudied by scholars and seems underappreciated by policymakers, especially when they consider their own reputations. Indeed, our impression from diplomatic histories and primary documents is that decision makers rarely think about their reputation for living up to promises when they withdraw from treaties or engage in behavior that others may see as renegeing. We know even less about how reputations for keeping promises form, how they relate to reputations for other characteristics, and how much weight they are given.

In practice, scholars of security studies remain almost myopically focused on resolve; the books we review here are no exception. This pattern persists despite important contributions from scholars of international political economy, including Michael Tomz, Julia Gray, and Raymond Hicks, which are outside of the scope of this article but are relevant nonetheless.<sup>66</sup> The myopic focus is understandable to the extent

<sup>61</sup> Brutger and Kertzer 2018.

<sup>62</sup> Sartori 2005; Crescenzi 2007.

<sup>63</sup> Dafoe, Renshon, and Huth 2014. Relatedly, George Downs and Michael Jones conclude that states have multiple reputations depending on the circumstances and the other actors involved in the interaction; see Downs and Jones 2002. For a discussion of how pariah states can seek to lose a bad reputation, see Clary 2020. Also see Lebow 1981; O'Neill 1999; Guisinger and Smith 2002; Gibler 2008; Miller 2012; Walter 2009; Peterson 2013; Henry 2020.

<sup>64</sup> Although see Davis 2000; Blankenship 2020.

<sup>65</sup> For an experimental study that touches on this topic but unfortunately does not address it head-on, see Heilman 1974, 323. But Heilman does find that "a reputation for matching deeds with words did not always guarantee that a new influence attempt would be credible. On the contrary such reputations (e.g., for fulfilling a threat) could impair credibility in new influence contexts (e.g., when making a promise). And, paradoxically, failing to fulfill commitments (e.g., welshing on a promise) could enhance future credibility (e.g., when making a threat)." The reason is that what was crucial was whether the other person was viewed positively or negatively. In the former case, the expectations were that she would fulfill promises but not threats; in the latter case it was the reverse.

<sup>66</sup> Tomz 2007b; Gray and Hicks 2014.

that resolve is viewed as essential for coercion, especially during crises. But we currently have no way to assess whether reputation for resolve has unique qualities that do not translate to other contexts. Jackson, to his credit, recognizes the tension between accumulating a reputation for resolve and honesty (pp. 17–18). All else equal, states would like to be known for meeting challenges and communicating frankly. But the choice to back down or to stand firm has different consequences for threat credibility and the likelihood of future challenges depending on the variety of reputation in question. Whereas a history of standing firm after having made a commitment to do so ought to garner a reputation for honesty and resolve no matter what, a history of backing down can show that a state is honest but not resolved, and a history of bluffing may indicate that the state is neither honest nor resolved. Signaling reputation can thus be orthogonal to a variety of substantive reputations.

Indeed, the discipline would benefit from additional research on the reputational consequences of bluffing or failing to live up to commitments to stand firm—particularly because we know relatively little about the effects of acting strongly in the absence of previous threats to do so. The obvious case is the American entry into the Korean War. The United States not only refrained from explicit threats to a fight if South Korea were invaded, but also withdrew its troops and declared the country to be outside the American defense perimeter. Stalin, Mao Zedong, and Kim Il-Sung were understandably surprised by the American action, but available documents do not tell us how this affected their image of the United States and their expectations of how it would behave in the future.

At minimum, the foregoing discussion suggests that different types of reputation can cut in different directions under some circumstances. The potential for cross-cutting effects raises a broader question about when these kinds of trade-offs emerge and how policymakers analyze them. For example, Crescenzi suggests that escalation to war is more likely in conflict-rich triadic settings when the antagonist has acquired a reputation for incompetence or belligerence in crises, perhaps even despite efforts to appear resolved (p. 116). Yet whether decision makers are even aware (beyond a basic understanding of the security dilemma) that they can overshoot on attempting to exhibit resolve and how policymakers address tensions in trying to cultivate varieties of reputations remain unclear. For example, if state A develops a reputation for complying with international agreements, it may signal that this state would be a contributor to new forms of cooperation, or conversely, state B might estimate that state A is a pushover and is likely to comply with

whatever terms it is offered. This dynamic offers another avenue for integrating psychological or dispositional perspectives, as these factors are almost certainly related to the types of reputation observing states and leaders prioritize in their own behavior and that of others.

Trade-offs may also arise with respect to the audience for different types of reputation. For instance, deception and secrecy are more or less accepted features of international politics. Nixon initiated the covert bombing of Cambodia despite the White House's public claims that the administration was seeking an end to the fighting in Southeast Asia. Not unrelatedly, Nixon sought to convince the Soviets that he was a "madman" to pressure the North Vietnamese into concessions.<sup>67</sup> Although international audiences may accept some degree of lying or posturing as a tool of statecraft, such actions may not ultimately sit well with domestic audiences.<sup>68</sup> Nixon's machinations in Cambodia earned him the ire of the American public and further diminished popular support for US involvement in the region. On the one hand, and to the extent that politics is a two-level game,<sup>69</sup> the perceived costs and benefits of developing a reputation for deception may vary according to which audience's opinion the leader values. On the other hand, leaders may value both audiences equally but may attempt to generate different reputations in the eyes of different audiences: Nixon's bombing campaign had the purpose of covertly signaling resolve to the North Vietnamese while avoiding charges of hypocrisy domestically as his administration attempted to wind down the conflict.<sup>70</sup>

## V. CONCLUSION

Reputations emerge, persist, and change according to observers' theories about how the world works and how best to explain others' behavior. New scholarship has embraced this relational perspective and delineated scope conditions for when reputation is most likely to matter, and has engaged in a range of methodological approaches along the way. But questions remain about the relationship between theories pitched at different units of analysis, the link between observer-level differences and associated judgments about credibility and reputation, the role of psychology in explaining how observers draw reputational inferences, and the importance of other sorts of reputations beyond that

<sup>67</sup> Varieties of madman theories are discussed in McManus 2019.

<sup>68</sup> For the claim that deception is almost always aimed at domestic, not foreign, audiences, see Mearsheimer 2011.

<sup>69</sup> Putnam 1988.

<sup>70</sup> Carson and Yarhi-Milo 2017.

for resolve. Although much progress has been made, our review makes clear that there remain many potentially fruitful avenues of inquiry in this research program.

To move the literature forward, scholars might consider a mix of research designs. Returning once more to the trade-off between depth and breadth, a major issue for future scholarship is the need to balance systematic studies that aim to deductively illustrate patterns across time and space with work that may evaluate only a few cases in the interest of inductive theory building. Achieving that balance is particularly important for developing literature that pursues new units of analysis, explores unit-level differences, leverages new insights from psychology and related fields, and probes various types of reputation. Most of these questions call for examining how the signaler and observer(s) act, perceive the other, and estimate how the other is perceiving them. But since the lion's share of existing work understandably focuses on only the signaler or observer,<sup>71</sup> a good deal of inductive theory building may be required to bridge these two perspectives prior to more general hypothesis testing.

Another potentially important dimension that scholars have left largely untouched is the connection between self-images and reputational concerns. We generally conceive of international politics as strategic interaction, but rarely do we address actors' degrees of self-consciousness when attempting to shape others' judgments about them. Although we know that signalers' attempts to project desired images often misfire because the observer interprets the behavior quite differently or fails to notice it at all,<sup>72</sup> we need to explore more thoroughly how this operates in the realm of reputation. Decades ago, Glen Snyder and Paul Diesing argued that while leaders were sure that their reputation in others' eyes strongly influenced others' behavior toward them, reputation played a much smaller role in determining their own behavior.<sup>73</sup> Yet there has been little research on this or related questions since. Are states' beliefs about their reputation in others' eyes generally accurate? Under what conditions are they more or less so? Do some actors, such as individuals, states, or regimes, have a better sense of how they are seen than others? For example, much of the work on how naming and shaming can bring leaders and states to pay greater respect to human rights rests on belief that these actors care not only about how others

<sup>71</sup> Although see Henry 2020 for a notable exception.

<sup>72</sup> For more discussion, Jervis 2017, chap. 5.

<sup>73</sup> Snyder and Diesing 1977; Tang 2005.

view them, but also on their own self-understandings and self-regard.<sup>74</sup> But if actors value their self-images (or various aspects of it) to different extents, there may be important scope conditions on the efficacy of naming and shaming, which makes it a key area for additional research.

We also need to consider how reputation, or reputations for various characteristics, relate to the concepts of honor, glory, status, and standing.<sup>75</sup> Here, reputations may be sought for reasons not easily characterized as instrumental and may drive us to look at the roots of behavior in culture and human nature, a once discredited topic that has been reinvigorated by a renewed study of biological factors.<sup>76</sup> It now seems clear, for example, that debates between rational choice and psychological perspectives offer a false dichotomy that ignores the emotional origins of preference orderings as well as the ways in which environmental and genetic factors interact in shaping political behavior. But more solid links between cutting-edge findings in biology and international relations research are only now being established.<sup>77</sup>

In sum, the books we review do an admirable job of addressing central questions in this research program: whether reputations form and the extent to which actors use others' past actions in anticipating future behavior. But as we turn toward future research, we outline several potential paths to move the literature forward and believe that advances in behavioral science as well as recent events, such as the current crisis over North Korea's nuclear program or the ongoing coronavirus pandemic, offer scholars innovative avenues to tackle both new and enduring questions in the study of credibility and reputation in international politics.

#### REFERENCES

- Blankenship, Brian. 2020. "Promises under Pressure: Statements of Reassurance in US Alliances." *International Studies Quarterly*. At <https://doi.org/10.1093/isq/sqaa071>.
- Brutger, Ryan, and Joshua D. Kertzer. 2018. "A Dispositional Theory of Reputation Costs." *International Organization* 72, no. 3: 693–724. At <https://doi.org/10.1017/S0020818318000188>.
- Byman, Daniel L., and Kenneth M. Pollack. 2001. "Let Us Now Praise Great Men: Bringing the Statesman Back In." *International Security* 25, no. 4: 107–46. At <https://doi.org/10.1162/01622880151091916>.
- Carson, Austin, and Keren Yarhi-Milo. 2017. "Covert Communication: The In-

<sup>74</sup> See, for example, Snyder 2020; Prakash and Potoski 2006; Lebovic and Voeten 2009.

<sup>75</sup> Kagan 1995; Lebow 2008; Paul, Larson, and Wohlforth 2014; Renshon 2017; Larson and Shevchenko 2019.

<sup>76</sup> Hatemi and McDermott 2011.

<sup>77</sup> Davis and McDermott 2020.

- telligibility and Credibility of Signaling in Secret." *Security Studies* 26, no. 1: 124–56. At <https://doi.org/10.1080/09636412.2017.1243921>.
- Clary, Matthew Q. 2020. *From Pariah to Phoenix: Improving a National Reputation from the Ashes of the Past*. Lanham, Md.: Rowman & Littlefield.
- Crescenzi, Mark J. C. 2007. "Reputation and Interstate Conflict." *American Journal of Political Science* 51, no. 2: 382–96. At <https://doi.org/10.1111/j.1540-5907.2007.00257.x>.
- Crescenzi, Mark J. C. 2018. *Of Friends and Foes: Reputation and Learning in International Politics*. New York, N.Y.: Oxford University Press.
- Dafoe, Allan, Jonathan Renshon, and Paul Huth. 2014. "Reputation and Status as Motives for War." *Annual Review of Political Science* 17: 371–93. At <https://doi.org/10.1146/annurev-polisci-071112-213421>.
- Davis, James W. 2000. *Threats and Promises: The Pursuit of International Influence*. Baltimore, Md: Johns Hopkins University Press.
- Davis, James W., ed. 2013. *Psychology, Strategy and Conflict: Perceptions of Insecurity in International Relations*. New York, N.Y.: Routledge.
- Davis, James W., and Rose McDermott. 2020. "The Past, Present and Future of Behavioral IR." *International Organization*: 1–31. At <https://doi.org/10.1017/S0020818320000272>.
- Downs, George W., and Michael A. Jones. 2002. "Reputation, Compliance, and International Law." *Journal of Legal Studies* 31, no. S1: S95–114. At <https://doi.org/10.1086/340405>.
- Elster, Jon. 1983. *Sour Grapes: Studies in Subversion of Rationality*. Cambridge, UK: Cambridge University Press.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88, no. 3: 577–92. At <https://doi.org/10.2307/2944796>.
- Gibler, Douglas M. 2008. "The Costs of Reneging: Reputation and Alliance Formation." *Journal of Conflict Resolution* 52, no. 3: 426–54. At <https://doi.org/10.1177/0022002707310003>.
- Goemans, Henk E., Kristian Skrede Gleditsch, and Giacomo Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research* 46, no. 2: 269–83. At <https://doi.org/10.1177/0022343308100719>.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Garden City, N.Y.: Doubleday & Co.
- Gray, Julia, and Raymond P. Hicks. 2014. "Reputations, Perceptions, and International Economic Agreements." *International Interactions* 40, no. 3: 325–49. At <https://doi.org/10.1080/03050629.2014.899227>.
- Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46, no. 2: 175–200. At <https://doi.org/10.1177/0022002702046002001>.
- Hafner-Burton, Emilie M., Stephan Haggard, David A. Lake, and David G. Victor. 2017. "The Behavioral Revolution and International Relations." *International Organization* 71, no. S1: S1–S31. At <https://doi.org/10.1017/S0020818316000400>.
- Harvey, Frank P., and John Mitton. 2016. *Fighting for Credibility: US Reputation and International Politics*. Toronto, Canada: University of Toronto Press.



- Hatemi, Peter K., and Rose McDermott. 2011. *Man Is by Nature a Political Animal: Evolution, Biology, and Politics*. Chicago, Ill.: University of Chicago Press.
- Heilman, Madeline E. 1974. "Threats and Promises: Reputational Consequences and Transfer of Credibility." *Journal of Experimental Social Psychology* 10, no. 4: 310–24. At [https://doi.org/10.1016/0022-1031\(74\)90028-6](https://doi.org/10.1016/0022-1031(74)90028-6).
- Henry, Iain D. 2020. "What Allies Want: Reconsidering Loyalty, Reliability, and Alliance Interdependence." *International Security* 44, no. 4: 45–83. At [https://doi.org/10.1162/isec\\_a\\_00375](https://doi.org/10.1162/isec_a_00375).
- Holmes, Marcus. 2018. *Face-to-Face Diplomacy: Social Neuroscience and International Relations*. Cambridge, UK: Cambridge University Press.
- Hopf, Ted. 1994. *Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965–1990*. Ann Arbor, Mich.: University of Michigan Press.
- Horowitz, Michael C., Allan C. Stam, and Cali M. Ellis. 2015. *Why Leaders Fight*. New York, N.Y.: Cambridge University Press.
- Huth, Paul K. 1997. "Reputations and Deterrence: A Theoretical and Empirical Assessment." *Security Studies* 7, no. 1: 72–99. At <https://doi.org/10.1080/09636419708429334>.
- Jackson, Van. 2016. *Rival Reputations: Coercion and Credibility in US–North Korea Relations*. Cambridge, UK: Cambridge University Press.
- Janis, Irving L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascos*. Boston, Mass.: Houghton Mifflin.
- Jervis, Robert. 1970. *The Logic of Images in International Relations*. Princeton, N.J.: Princeton University Press.
- Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press.
- Jervis, Robert. 1998. *System Effects: Complexity in Political and Social Life*. Princeton, N.J.: Princeton University Press.
- Jervis, Robert. 2013. "Do Leaders Matter and How Would We Know?" *Security Studies* 22, no. 2: 153–79. At <https://doi.org/10.1080/09636412.2013.786909>.
- Jervis, Robert. 2017. *How Statesmen Think: The Psychology of International Politics*. Princeton, N.J.: Princeton University Press.
- Jervis, Robert, and Jack L. Snyder. 1991. *Dominoes and Bandwagons: Strategic Beliefs and Great Power Competition in the Eurasian Rimland*. New York, N.Y.: Oxford University Press.
- Joll, James. 1968. *1914: The Unspoken Assumptions: An Inaugural Lecture Delivered 25 April 1968*. London, UK: London School of Economics and Political Science and Weidenfeld & Nicolson.
- Kagan, Donald. 1995. *On the Origins of War: And the Preservation of Peace*. New York, N.Y.: Anchor Books.
- Kertzer, Joshua D. 2016. *Resolve in International Politics*. Princeton, N.J.: Princeton University Press.
- Kertzer, Joshua D. Forthcoming. "Re-Assessing Elite-Public Gaps in Political Behavior." *American Journal of Political Science*. At [https://people.fas.harvard.edu/~jkertzer/Research\\_files/Elite-Public-Gaps-Web.pdf](https://people.fas.harvard.edu/~jkertzer/Research_files/Elite-Public-Gaps-Web.pdf), accessed October 28, 2020.
- Kertzer, Joshua D., Brian C. Rathbun, and Nina Srinivasan Rathbun. 2020. "The

- Price of Peace: Motivated Reasoning and Costly Signaling in International Relations." *International Organization* 74, no. 1: 95–118. At <https://doi.org/10.1017/S0020818319000328>.
- Kertzer, Joshua D., Jonathan Renshon, and Keren Yarhi-Milo. 2019. "How Do Observers Assess Resolve?" *British Journal of Political Science*: 1–23. At <https://doi.org/10.1017/S0007123418000595>.
- Kertzer, Joshua D., and Dustin Tingley. 2018. "Political Psychology in International Relations: Beyond the Paradigms." *Annual Review of Political Science* 21: 319–39. At <https://doi.org/10.1146/annurev-polisci-041916-020042>.
- Khong, Yuen Foong. 1992. *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965*. Princeton, N.J.: Princeton University Press.
- Kuzmanovic, Bojana, and Lionel Rigoux. 2017. "Valence-Dependent Belief Updating: Computational Validation." *Frontiers in Psychology* 8. At <https://doi.org/10.3389/fpsyg.2017.01087>.
- Lamb, Christopher J. 2018. "The Mayaguez Crisis: Correcting 30 Years of Scholarship." *Political Science Quarterly* 133, no. 1: 35–76. At <https://doi.org/10.1002/polq.12736>.
- Larson, Deborah Welch, and Alexei Shevchenko. 2019. *Quest for Status: Chinese and Russian Foreign Policy*. New Haven, Conn.: Yale University Press.
- Lebovic, James H., and Erik Voeten. 2009. "The Cost of Shame: International Organizations and Foreign Aid in the Punishing of Human Rights Violators." *Journal of Peace Research* 46, no. 1: 79–97. At <https://www.jstor.org/stable/27640800>.
- Lebow, Richard Ned. 1981. *Between Peace and War: The Nature of International Crisis*. Baltimore, Md.: Johns Hopkins University Press.
- Lebow, Richard Ned. 2008. *A Cultural Theory of International Relations*. Cambridge, UK: Cambridge University Press.
- Levendusky, Matthew S., and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *Journal of Politics* 74, no. 2: 323–38. At <https://doi.org/10.1017/S002238161100154X>.
- Levy, Jack S., Michael K. McKoy, Paul Poast, and Geoffrey P. R. Wallace. 2015. "Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory." *American Journal of Political Science* 59, no. 4: 988–1001. At <https://doi.org/10.1111/ajps.12197>.
- Lin-Greenberg, Erik. 2019. "Backing Up, Not Backing down: Mitigating Audience Costs through Policy Substitution." *Journal of Peace Research* 56, no. 4: 559–74. At <https://doi.org/10.1177/0022343319832641>.
- Logevall, Fredrik. 1999. *Choosing War: The Lost Chance for Peace and the Escalation of War in Vietnam*. Berkeley, Calif.: University of California Press.
- Lupton, Danielle L. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca, N.Y.: Cornell University Press.
- Marinov, Nikolay. 2005. "Do Economic Sanctions Destabilize Country Leaders?" *American Journal of Political Science* 49, no. 3: 564–76. At <https://doi.org/10.1111/j.1540-5907.2005.00142.x>.
- Markwica, Robin. 2018. *Emotional Choices: How the Logic of Affect Shapes Coercive Diplomacy*. New York, N.Y.: Oxford University Press.
- McGillivray, Fiona, and Allan C. Stam. 2004. "Political Institutions, Coercive Diplomacy, and the Duration of Economic Sanctions." *Journal of Conflict Resolution* 48, no. 2: 154–72. At <https://doi.org/10.1177/0022002703262858>.

- McManus, Roseanne W. 2017. *Statements of Resolve: Achieving Coercive Credibility in International Conflict*. Cambridge, UK: Cambridge University Press.
- McManus, Roseanne W. 2019. "Revisiting the Madman Theory: Evaluating the Impact of Different Forms of Perceived Madness in Coercive Bargaining." *Security Studies* 28, no. 5: 976–1009. At <https://doi.org/10.1080/09636412.2019.1662482>.
- Mearsheimer, John J. 2011. *Why Leaders Lie: The Truth About Lying in International Politics*. New York, N.Y.: Oxford University Press.
- Mercer, Jonathan. 1996. *Reputation and International Politics*. Ithaca, N.Y.: Cornell University Press.
- Mercer, Jonathan. 2010. "Emotional Beliefs." *International Organization* 64, no. 1: 1–31. At <https://doi.org/10.1017/S0020818309990221>.
- Mercer, Jonathan. 2013a. "Emotion and Strategy in the Korean War." *International Organization* 67, no. 2: 221–52. At <https://doi.org/10.1017/S0020818313000015>.
- Mercer, Jonathan. 2013b. "Rational Signaling Revisited." In James W. Davis, ed., *Psychology, Strategy and Conflict: Perceptions of Insecurity in International Relations*. New York, N.Y.: Routledge: 64–81.
- Miller, Gregory D. 2012. *The Shadow of the Past: Reputation and Military Alliances Before the First World War*. Ithaca, N.Y.: Cornell University Press.
- Mintz, Alex. 2004. "How Do Leaders Make Decisions? A Poliheuristic Perspective." *Journal of Conflict Resolution* 48, no. 1: 3–13. At <https://doi.org/10.1177/0022002703261056>.
- O'Neill, Barry. 1999. *Honor, Symbols, and War*. Ann Arbor, Mich.: University of Michigan Press.
- Paul, T. V., Deborah Welch Larson, and William C. Wohlforth. 2014. *Status in World Politics*. New York, N.Y.: Cambridge University Press.
- Peterson, Timothy M. 2013. "Sending a Message: The Reputation Effect of US Sanction Threat Behavior." *International Studies Quarterly* 57, no. 4: 672–82. At <https://doi.org/10.1111/isqu.12017>.
- Prakash, Aseem, and Matthew Potoski. 2006. "Racing to the Bottom? Trade, Environmental Governance, and ISO 14001." *American Journal of Political Science* 50, no. 2: 350–64. At <https://www.jstor.org/stable/3694277>.
- Press, Daryl Grayson. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. Ithaca, N.Y.: Cornell University Press.
- Putnam, Robert D. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42, no. 3: 427–60. At <https://doi.org/10.1017/S0020818300027697>.
- Renshon, Jonathan. 2017. *Fighting for Status: Hierarchy and Conflict in World Politics*. Princeton, N.J.: Princeton University Press.
- Renshon, Jonathan, Allan Dafoe, and Paul Huth. 2018. "Leader Influence and Reputation Formation in World Politics." *American Journal of Political Science* 62, no. 2: 325–39. At <https://doi.org/10.1111/ajps.12335>.
- Samanez-Larkin, Gregory R., and Brian Knutson. 2015. "Decision Making in the Ageing Brain: Changes in Affective and Motivational Circuits." *Nature Reviews Neuroscience* 16: 278–89. At <https://doi.org/10.1038/nrn3917>.
- Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56, no. 1: 121–49. At <https://doi.org/10.1162/002081802753485151>.

- Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton, N.J.: Princeton University Press.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, Conn.: Yale University Press.
- Schramm, Madison, and Alexandra Stark. 2020. "Peacemakers or Iron Ladies? A Cross-National Study of Gender and International Conflict." *Security Studies* 29, no. 3: 515–45. At <https://doi.org/10.1080/09636412.2020.1763450>.
- Schwartz, Joshua A., and Christopher W. Blair. 2020. "Do Women Make More Credible Threats? Gender Stereotypes, Audience Costs, and Crisis Bargaining." *International Organization*: 1–24. At <https://doi.org/10.1017/S0020818320000223>.
- Sharot, Tali, and Neil Garrett. 2016. "Forming Beliefs: Why Valence Matters." *Trends in Cognitive Sciences* 20, no. 1: 25–33. At <https://doi.org/10.1016/j.tics.2015.11.002>.
- Shultz, George P. 1996. *Turmoil and Triumph: Diplomacy, Power, and the Victory of the American Deal*. New York, N.Y.: Simon and Schuster.
- Snyder, Glenn Herald, and Paul Diesing. 1977. *Conflict among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, N.J.: Princeton University Press.
- Snyder, Jack. 2020. "Backlash against Human Rights Shaming: Emotions in Groups." *International Theory* 12, no. 1: 109–32. At <https://doi.org/10.1017/S1752971919000216>.
- Sullivan, Andy. 2013. "Most in U.S. Still Oppose Syria Strike as Obama Seeks Action." *Reuters*, August 30, 2013. At <https://www.reuters.com/article/us-syria-crisis-usa-poll-idUSBRE97T0UO20130830>, accessed October 28, 2020.
- Tang, Shiping. 2005. "Reputation, Cult of Reputation, and International Conflict." *Security Studies* 14, no. 1: 34–62. At <https://doi.org/10.1080/09636410591001474>.
- Tomz, Michael. 2007a. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61, no. 4: 821–40. At <https://www.jstor.org/stable/4498169>.
- Tomz, Michael. 2007b. *Reputation and International Cooperation: Sovereign Debt across Three Centuries*. Princeton, N.J.: Princeton University Press.
- Walter, Barbara F. 2009. *Reputation and Civil War: Why Separatist Conflicts Are So Violent*. New York, N.Y.: Cambridge University Press.
- Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62, no. 1: 35–64. At <https://www.jstor.org/stable/40071874>.
- Weisiger, Alex, and Keren Yarhi-Milo. 2015. "Revisiting Reputation: How Past Actions Matter in International Politics." *International Organization* 69, no. 2: 473–95. At <https://doi.org/10.1017/S0020818314000393>.
- Wheeler, Nicholas J. 2018. *Trusting Enemies: Interpersonal Relationships in International Conflict*. Oxford, UK: Oxford University Press.
- Wolford, Scott. 2007. "The Turnover Trap: New Leaders, Reputation, and International Conflict." *American Journal of Political Science* 51, no. 4: 772–88. At <https://doi.org/10.1111/j.1540-5907.2007.00280.x>.
- Wu, Cathy Xuanxuan, and Scott Wolford. 2018. "Leaders, States, and Reputations." *Journal of Conflict Resolution* 62, no. 10: 2087–117. At <https://doi.org/10.1177/0022002718786001>.

- Yarhi-Milo, Keren. 2014. *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations*. Princeton, N.J.: Princeton University Press.
- Yarhi-Milo, Keren. 2018. *Who Fights for Reputation: The Psychology of Leaders in International Conflict*. Princeton, N.J.: Princeton University Press.
- Yarhi-Milo, Keren, Joshua D. Kertzer, and Jonathan Renshon. 2018. "Tying Hands, Sinking Costs, and Leader Attributes." *Journal of Conflict Resolution* 62, no. 10: 2150–79. At <https://doi.org/10.1177/0022002718785693>.
- Zhang, Ketian. 2019. "Cautious Bully: Reputation, Resolve, and Beijing's Use of Coercion in the South China Sea." *International Security* 44, no. 1: 117–59. At [https://doi.org/10.1162/isec\\_a\\_00354](https://doi.org/10.1162/isec_a_00354).

### AUTHORS

ROBERT JERVIS is a professor of international politics in the department of political science and the School of International and Public Affairs at Columbia University. He specializes in international politics with a focus on security policy, decision-making, and theories of conflict and cooperation. A prolific author, his books include *How Statesmen Think* (2017) and *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War* (2010). He can be reached at [rlj1@columbia.edu](mailto:rlj1@columbia.edu).

KEREN YARHI-MILO is a professor of war and peace studies in the department of political science and the School of International and Public Affairs at Columbia University. Her research and teaching focus on international relations and foreign policy, with a particular specialization in international security, including foreign policy decision-making, interstate communication and crisis bargaining, psychology, intelligence, and US foreign policy in the Middle East. Her books include *Who Fights Reputation? The Psychology of Leaders in International Conflict* (2018). She can be reached at [ky218@columbia.edu](mailto:ky218@columbia.edu).

DON CASLER is a doctoral candidate in political science at Columbia University. He studies topics at the intersection of international security and international political economy, and his research interests include foreign policy decision-making, political psychology, and public opinion. He can be reached at [dpc2128@columbia.edu](mailto:dpc2128@columbia.edu).

### ACKNOWLEDGMENTS

The authors are grateful to Josh Kertzer, the editors of *World Politics*, and three anonymous reviewers for their helpful suggestions. Inbar Peer provided excellent research assistance.

### KEY WORDS

credibility, foreign policy, leaders, perception, psychology, reputation, signaling