

Reputations in the Wild: Reputation Costs from the Syrian and Crimean Crises Among Leaders and Citizens

Joshua D. Kertzer¹, Jonathan Renshon² & Keren Yarhi-Milo³

Abstract: While an older conventional wisdom held that reputation was “one of the few things worth fighting for,” later work suggested that reputations either did not matter or possibly did not exist at all. We argue that much of the recent “debate” over reputation has been hamstrung by scholars asking and answering entirely different questions about reputations. We introduce a new conceptual framework that divides reputational inferences into the *evaluation* phase (in which observers update their beliefs about type/attributes) and a *diagnostic* phase in which judgments are made about present and future behavior. We then introduce a new type of evidence into this debate, presenting results from original surveys fielded on four samples (including Israeli decision-makers, Israeli publics, and American IR scholars) to study the *evaluation* phase in which reputation costs were incurred by various actors as a result of the Russian invasion of Crimea and the Syrian civil war. Our results suggest that reputation costs exist, are worse in cases of failed threats, attach to both leaders and the countries they represent, and are perceived similarly by elite leaders and ordinary citizens. Finally, our studies revealed the existence of a “home-away gap”: compared to foreign observers, American IR scholars seem to underestimate the magnitude of reputation costs the US has incurred by backing down on threats.

Word count: 8,000 words (including bibliography)

¹Professor of Government, Department of Government, Harvard University. Email: jkertzer@gov.harvard.edu. Web: <http://people.fas.harvard.edu/~jkertzer/>

²Professor, Department of Political Science, University of Wisconsin-Madison. Email: renshon@wisc.edu. Web: <http://jonathanrenshon.net>

³Arnold A. Saltzman Professor of War and Peace Studies, Department of Political Science & The School of International and Public Affairs Columbia University. Email: ky218@columbia.edu. Web: <https://polisci.columbia.edu/content/keren-yarhi-milo>.

One of the most central debates in U.S. foreign policy concerns the role of reputation (beliefs about persistent attributes or “type”). While an older tradition saw a reputation for resolve as “one of the few things worth fighting for” (Schelling, 1966, 124), a prominent dissenting tradition argued that foreign observers pay less attention to past actions than early deterrence theorists believed and that reputations may not be updated appropriately or even form at all (Hopf, 1994; Mercer, 1996; Tang, 2005; Press, 2005). This “reputation cynic” argument was sufficiently well-entrenched that the *New York Times* summarized the state of IR scholarship by noting that arguments about the importance of reputation were “baseless” and had “been extensively debunked” (Fisher, 2017). While we see the reputation “cynics” as being overwhelmed both in number and by the available evidence, the debate has persisted: even newer research in this area has framed significant parts of their contribution as being about “whether reputations matter” (e.g., Lupton, 2020, 2; Harvey and Mitton, 2017, 4)

In this research note, we follow the guidance of two prominent reviews in focusing our inquiry on “when, how and why” reputations matter (Dafoe, Renshon and Huth, 2014, 372), with specific attention paid to “unpacking” the variation in perceptions across actors and audiences noted by Jervis, Yarhi-Milo and Casler (2021, 168). We make two principal contributions. First, we argue that what has seemed like a fiery debate about reputations has resulted from theories that ask and answer significantly different questions. Our conceptual contribution is thus to decompose reputational inferences into two phases: the *evaluation* phase in which behavior is observed by an audience, who either update their beliefs about the actor’s type and attributes or not, and the *diagnostic* phase, in which judgments are made about present and future behavior. The former addresses the question of whether observers draw inferences about reputation in response to an actor’s past actions in the manner that classical theories of reputation assume, while the latter asks: how much weight do observers attach to reputational inferences when planning present behavior? Whether reputations “matter,” for example, is a question about the weight accorded reputations—alongside other factors—in the calculation of what others are likely to do in the future (our *diagnostic* phase; e.g., Weisiger and Yarhi-Milo, 2015; Wu and Wolford, 2018), while questions about how reputations form and are updated (e.g., Renshon, Dafoe and Huth, 2018) are questions about the *evaluation* phase.

Our empirical contribution hones in on what we call “reputation costs”: the effect of an actor’s behavior as judged by varying audiences, an estimand that implicates the *evaluation* phase

of reputational inferences. Our focus is on two sets of questions related to reputation costs: the nature and existence of reputation costs and variation in those costs both across targets of inference (different actors) and across audiences. We explore real-world reputation costs using four surveys, innovating in three key respects. First, we match our method to our concept of interest, turning to surveys to study individual beliefs in an area thus far dominated by formal theory, case studies, experiments and large- N analysis. Previous research at the micro level has often used experiments designed around hypothetical actors in abstract vignettes and, while such research is not inherently flawed in the ways that skeptics sometime worry about (Brutger et al., 2022), we argue that there are significant benefits to designs that leverage real-world events and actors and allow us to calibrate the magnitude of effects found in highly abstract experiments where observers have no real priors about the context.

Our second (related) innovation is to focus on two highly salient cases: the conflict in Crimea, where Russian forces “reconnected” with the (formerly Ukrainian) territory in late 2014 over the objections of the United States, and the conflict in Syria, where the Obama administration repeatedly declared the use of chemical weapons by the Assad regime to be a “red line” that would bring the United States into the conflict, only to subsequently back down. Since one problem in the study of reputation is that behavior like this is observationally rare—if reputation costs are real, leaders should do their best to avoid them—these cases thus provide important and unusual insight into the scale of reputational costs “in the wild.” Finally, we field our surveys on a sample of elite decision-makers—current and former members of the Israeli Knesset—alongside a sample of IR scholars in the United States, and two nationally representative samples of the Israeli public. These samples implicate the domain-specific knowledge of two types of elites (Kertzer and Renshon, 2022) and allow us to compare reputation costs suffered to the U.S. and Russia across audiences of both domestic and foreign observers, as well as the general public and elites.

We find evidence that speaks to both general patterns in the *evaluation phase* of reputational inferences, and to specific questions related to how the reputations of key world leaders were affected—in the eyes of varying audiences—by their behavior. We find that President Obama and the United States suffered substantial reputation costs in both crises, while Russia’s reputation appears to have been *enhanced* following its actions in Syria. We also find that the reputational damage for the United States was more stark in the case of Syria than Crimea—consistent with the predictions of audience cost theory. In contrast to these models, however, our results focus attention on the

international ramifications of failed threats rather than their domestic electoral consequences. Finally, we find that IR scholars at home perceive significantly smaller reputation costs compared to observers abroad, suggesting that IR scholars may be underestimating reputational inferences.

1 The Rise and Fall of Reputations in IR

Does reputation for resolve matter in international relations? Thomas Schelling led a “first wave” of scholars who saw reputations as critical to IR, famously arguing that the United States had sacrificed thirty-thousand dead in the Korean War, but that doing so was “undoubtedly worth it” since “Soviet expectations about the behavior of the United States are one of the most valuable assets we possess in world affairs” (Schelling, 1966, 125-126). Like much of his other work, Schelling’s clear, direct and powerful insights gained widespread acceptance within IR, setting a research agenda for several decades following his original work.

As the Cold War drew to a close, a group of “reputation critics” began to challenge the claim that reputations for resolve critically affected international politics, or even existed at all. Ted Hopf (1994, 117) showed that, as intuitively compelling as Schelling’s model of deterrence and reputation was, there was little evidence of it at work in the Cold War: no matter what the U.S. did, it seemed unable to shake its reputation for resolve. This work, along with subsequent research (e.g., Mercer, 1996; Press, 2005), challenged some of Schelling’s key assumptions, such as the notion that commitments were interdependent or even that reputations could form in the first place. Tang (2005) provides the strongest statement of the “reputation critics,” arguing that there is a “cult” of reputation among political leaders, who expend resources to chase something that doesn’t exist. This conception of reputation has reached the highest levels: Barack Obama, in discussing notions of credibility and reputation, argued: “Look, this theory is so easily disposed of that I’m always puzzled by how people make the argument.” Jeffrey Goldberg, who conducted that interview, summarized Obama’s views by noting that he “generally believes that the Washington foreign-policy establishment, which he secretly disdains, makes a fetish of ‘credibility’” (Goldberg, 2016).

Reputation “cynics” spurred pushback (and a third wave of research) against the claim that reputations do not matter (Harvey and Mitton, 2017; Lupton, 2020). Using large-N observational data, Weisiger and Yarhi-Milo (2015) have shown the consequences of reputations: countries that backed down in previous disputes are significantly more likely to be challenged again, while states that

stood firm are less likely to be challenged in the future (see also [Clare and Danilovic, 2010](#); [Kertzer, Renshon and Yarhi-Milo, 2021](#)). This third wave has not only branched off into a vibrant strain of work on audience costs—which relies heavily on the assumption that reputations for resolve exist, that publics care about maintaining them, and that irresolute behavior harms them—but also disaggregated reputation among states and leaders ([Tomz, 2007](#); [Weisiger and Yarhi-Milo, 2015](#); [Casler and Clark, 2021](#)) and shed light on how reputations form among observers of conflicts ([Crescenzi, 2018](#)).

1.1 Decomposing Reputational Inferences: Evaluation and Diagnostic Phases

One important source of the disagreement noted above about reputations in IR, we argue, stems from scholars (implicitly) asking different questions. For reputations to “matter” in IR, two conditions have to be met. First, observers must form or change beliefs about an actor in response to behavior or actions that they observe. We refer to this as the *evaluation* phase since behavior on the international stage is evaluated by observers and incorporated into their view of that actor, their type and their attributes; in short, their reputation. Second, observers must then draw upon reputational considerations when considering what other actors will do in the present/future. We refer to this latter concept as the *diagnostic* phase, since it asks whether (and how much) diagnostic weight in an analysis is placed on reputation compared to other variables like the intrinsic interests at stake. Reputations existing and shifting in response to events is thus a necessary but not sufficient condition for the diagnostic phase: without the formation of beliefs about reputation that are sensitive to updating based on observed behavior, reputational considerations cannot “matter” in any decision-making process.

Various critiques of reputation emphasize different phases of this causal chain. [Tang \(2005, 50\)](#) largely focuses on the evaluation phase, questioning whether actors draw reputational inferences the way our theories say they should. [Press \(2005, 1\)](#), on the other hand, largely focuses on the diagnostic phase, arguing that “when assessing credibility during crises, leaders focus on the ‘here and now,’ not on their adversary’s past behavior.” Similarly, advocates for reputations “mattering” also implicate different phases. [Renshon, Dafoe and Huth \(2018\)](#) target the evaluation phase in their theory of how international behavior will differentially affect the reputations of leaders and states,

while work by [Lupton \(2018\)](#) and [Weisiger and Yarhi-Milo \(2015\)](#) implicates the diagnostic phase by examining the effect of past actions on expectations of future behavior.

We tackle the *evaluation* phase of reputations head-on, testing whether and how reputations form and shift in response to two of the defining foreign policy crises of the Obama administration: the 2014 Russian annexation of Crimea, and the humanitarian crisis in the Syrian Civil War. We focus on the evaluation phase for two reasons. First, reputation costs have received relatively little empirical focus compared to questions related to the “diagnostic” phase. Much of the large-N literature on reputation, because it cannot directly study observers’ beliefs, concentrates instead on observable implications like crisis initiation—rendering it a better test of how and whether reputational considerations matter in affecting future behavior than whether reputations form and update the way our theories suggest they should.

Second, reputation costs offers a particularly significant challenge to existing theory: both studies of reputation in IR and the wider literature on judgment and decision-making have shown that individuals deviate from canonical models of updating ([Rathbun, 2014](#)); that they are “prisoners of their preconceptions” ([Tetlock, 1999](#)), resistant to change and driven by motivated biases (as prevalent among elites as the mass public, [Kertzer, 2022](#)). Even work largely supportive of reputations “mattering” suggests cause for concern about the initial formation and updating of reputation ([Brutger and Kertzer, 2018](#)). In contrast, whether reputations “matter” is in some ways overdetermined: although it would be irrational for observers to assume that past actions are perfectly predictive of future behavior given that no two situations are perfectly alike, it would be similarly irrational for observers to assume that past actions have no diagnostic value whatsoever ([Harvey and Mitton, 2017](#)). Most other disciplines—from economics, to sociology, to evolutionary biology—emphasize how conditions of uncertainty prompt actors to turn to past actions for information, such that it would be surprising if reputational diagnoses materialized in every other domain of human life apart from international politics.

1.2 Hypotheses on Reputations in the Evaluation Phase

Our empirical focus is on how reputations form and shift in response to actors’ behavior on the international stage: i.e., the *reputation costs* that actors pay for their actions. Our first set of hypotheses (depicted in Figure 1 along with the specific empirical strategy for each hypothesis) concerns the existence and nature of reputation costs: whether they exist; if they do, whether they

persist over time and whether (as the literature suggests) inaction following “red line” threats is worse than inaction alone.

H1 follows from the debate about whether reputations form in the first place; i.e, whether past behavior changes evaluations about a country’s underlying attributes or type. This is a pre-condition for both updating of reputations and for reputations to “matter” in evaluating future behavior, yet there are strong arguments against it (Tang, 2005) and numerous empirical works which call into question this basic premise (e.g., Hopf, 1994; Mercer, 1996; Shannon and Dennis, 2007). H2 follows the formation of reputations by proposing that—drawing from cognitive and psychological work on beliefs—reputational inferences are sticky and change only slowly (Crescenzi, 2018). However, it’s also plausible that as time goes by, assessments based on past actions will become less “available” as a heuristic or that past events will seem less relevant as contextual differences multiply over time. H3 concerns the *magnitude* of reputational costs in response to different behaviors. In particular, this hypothesis follows directly from the audience costs literature, which hinges on the notion that public threats constitute a particularly costly signal of resolve because backing down from those threats would significantly hurt the state’s reputation. Therefore, backing down from a threat should be associated with a steeper reputational cost compared to inaction that occurs without a threat.

Our second set of hypotheses follows from Jervis, Yarhi-Milo and Casler’s (2021) exhortation to investigate variation in reputational inferences and perceptions across actors and audiences. We do that in two ways: by examining variation in reputation costs across the *targets* of reputational inference (different actors) as well as across audiences. H4—on leader/state reputations—follows from work by Weisiger and Yarhi-Milo (2015) as well as Brutger and Kertzer (2018), which demonstrates that reputations in general adhere to both the leader and the state. At the same time, recent research from Renshon, Dafoe and Huth (2018), McManus (2018), and Lupton (2020) argues that reputation will adhere more to leaders in the domain of international security, especially when the leader stakes his personal reputation on a public threat. As a result, leader-specific reputations can dominate state-specific ones under particular conditions.

Finally, H5 tests observable implications about reputation costs varying across audiences. While studies of audience costs have focused on the domestic public’s concern for their state’s reputation, we investigate whether those concerns are justified by asking if states actually pay reputation costs among international audiences. Building on work by Mercer (1996), we postulate that motivated reasoning and attribution biases make domestic observers less likely than foreign audiences to view

	Hypothesis	Prediction	General Empirical Strategy
The nature and existence of reputation costs	H1: Existence of reputation costs	Observers form beliefs about another country's reputation based on its actions in an international crises.	Survey questions about how reputations have been affected by how actors have handled events in Syria/Crimea
	H2: Stability of reputation costs	Observers' beliefs about another state's reputation linger, although their strength may decay over time.	Comparison of survey responses at time t and second wave fielded at $t + 105$ days
	H3: Magnitude of reputation costs	Observers are more likely to downgrade another state's reputation when its leader backs down from a public threat compared to when no public threat has been issued	Contrast between Crimea/Syria responses
Variation in Reputation Costs Across Targets and Audiences	H4: Reputation costs across leaders and states	Observers will assign reputations to both states and leaders, but when a public threat is issued, reputational assessments will adhere more to the leader than to the state.	Contrast between individual leaders (Obama/Putin) and country (U.S./Russia) responses
	H5: Reputation costs as judged by varying audiences (foreign/domestic and elite/public)	(a) Relative to the domestic audience, foreign audiences will downgrade the other country's reputation more heavily when it backs down, while we expect (b) broad congruence between elites/citizens .	Contrast between foreign audiences (Israeli citizens and Knesset members) and domestic (U.S. academics) and between Israeli citizens and leaders
*H1, H3, H4 and H5 leverage all four samples; H2 relies on the two samples of Israeli citizens.			

Figure 1: Hypotheses

their own country backing down as a signal of weakness. In contrast, international audiences should be less predisposed to display motivated reasoning and attribution biases, which should lead them to downgrade the reputation of the other country after it backs down. Importantly, we do not argue that domestic audiences will fail to punish their leaders for backing down, but rather that we are likely to observe a “home-away” gap in the magnitude of these reputational costs. With respect to differences between elites and citizens, we expect—in line with work by [Renshon, Yarhi-Milo and Kertzer \(2023\)](#) and meta-level findings by [Kertzer \(2022\)](#)—broad congruence between the leaders and their publics.

2 Empirical Context, Data and Methods

We study reputational inferences from two of the defining foreign policy crises of the Obama administration: the Russian annexation of Crimea in the spring of 2014, and the humanitarian crisis in Syria, particularly in the aftermath of Obama’s declaration that the use of chemical weapons by the Assad regime would constitute a “red line” that would lead to intervention by the international community. To motivate our context, we use an original automated content analysis of over 87,000 newspaper articles on the two crises in Figure 2 to show that reputational rhetoric loomed large in each crisis *even at that time*. Figure 3 plots the main developments in each crisis.¹ In the Syrian

¹Timeline of events in Syria from: [BBC News Syria Profile](#) and [POLITICO Timeline: U.S. Approach to the Syrian Civil War](#). Information on the Crimean conflict comes: [BBC News Ukraine Crisis Timeline](#) and [CNN Ukraine: Everything You Need to Know About How We Got Here](#).

crisis, Barack Obama issued his famous “red line” threat in August of 2012 following the tightening of U.S./EU sanctions and concerns that Assad might use non-traditional weapons to suppress the uprising:

We have been very clear to the Assad regime, but also to other players on the ground, that a red line for us is we start seeing a whole bunch of chemical weapons moving around or being utilized. That would change my calculus.

Despite the attention paid to the so-called “red-line” statement, the U.S. ultimately did not intervene significantly in Syria, Assad remained in power and the uprising erupted into a full-blown civil war. In stark contrast, the Crimea timeline shows a very different type of conflict. In this case, Obama “urged” Putin to withdraw troops and enacted sanctions (along with the EU, at times), but avoided the stark language of threats he had utilized in the Syrian crisis. In many ways, however, the outcome was similar from the U.S. perspective: a failure to accomplish its policy goals and a looming concern that it would pay a price for its inaction in the coin of credibility.

Figure 2: Reputational language in media coverage of Syria and Crimea crises

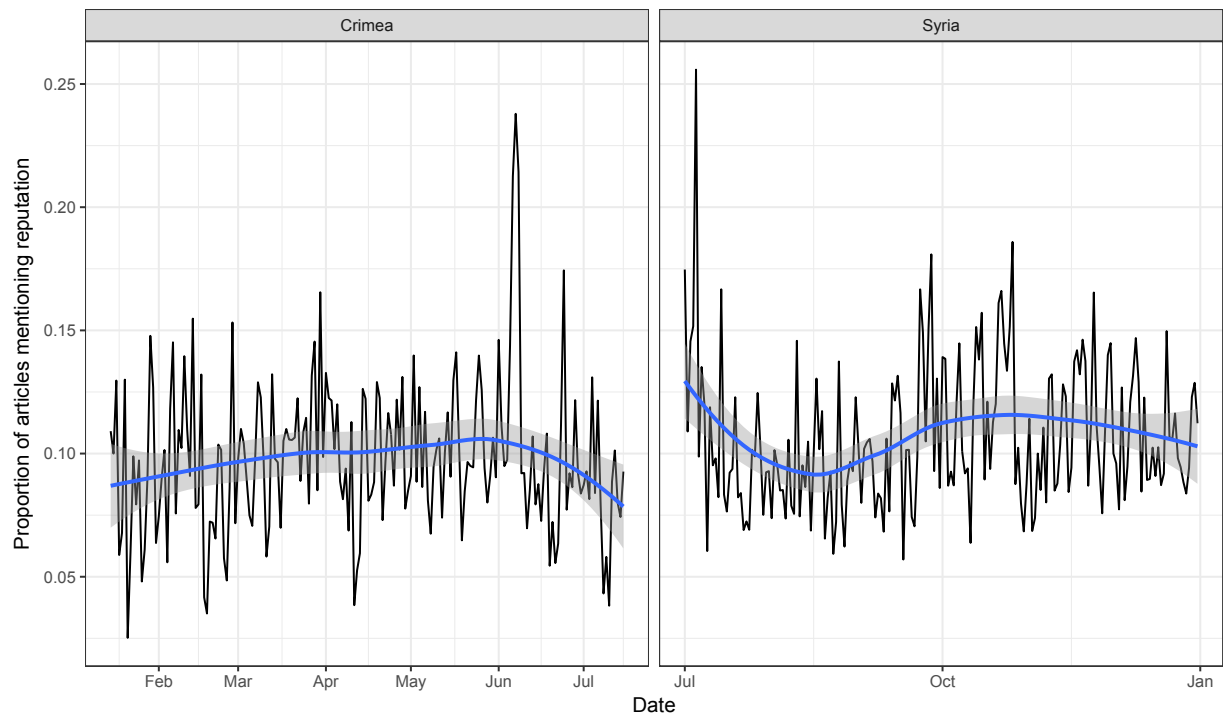


Figure 2 estimates the proportion of reputational language in daily newspaper coverage of the crisis in Crimea (from January 15-July 15, 2014) and Syria (from July 1-December 31, 2012), superimposed by a loess smoother (in blue). See details of coding in Appendix.

Syria	Crimea
<ul style="list-style-type: none"> • <i>March 2011</i>: Uprising begins; Assad announces conciliatory measures (e.g., lifts 48-year old state of emergency). • <i>May 2011</i>: Tanks enter suburbs of Damascus in an effort to crush anti-regime protests. US and EU tighten sanctions. • <i>August 2011</i>: President Barack Obama (joined by Germany, France and UK) formally calls on Assad to step down “for the sake of the Syrian people” and issues new sanctions on his regime. • <i>June 2011</i>: IAEA reports Syria to UN Security Council over covert nuclear program • <i>August 20, 2012</i>: In remarks to reporters at the White House, Obama warns Assad’s regime not to use chemical weapons, describing it as a “red line” for his administration that could prompt the United States to intervene militarily. • <i>August 24, 2012</i>: The USS John C. Stennis and its strike group prepares to deploy to the Middle East. • <i>August 29, 2012</i>: Senator John McCain McCain says Obama’s inaction in Syria has led to a “savage and unfair fight” by the regime of Bashar Assad. • <i>August 31, 2012</i>: A leading member of the Free Syrian Army says that, contrary to media reports, they have received no operational support from the United States or any other state. • <i>August 21, 2013</i>: Hundreds of people suffocate in rebel-held suburbs of the Syrian capital. UN investigators visit the sites and determine that ground-to-ground missiles loaded with sarin were fired on civilian areas while residents slept. The US and others blame the Syrian government, the only party known to have sarin gas. • <i>September 27, 2013</i>: UN Security Council orders Syria to account for and destroy its chemical weapons stockpile; threatens to authorize the use of force in the event of non-compliance. • <i>January-February 2014</i>: UN-brokered peace talks in Geneva fail, largely because Syrian authorities refuse to discuss a transitional government. • <i>March 2014</i>: Syrian Army/Hezbollah forces recapture Yabroud, the last rebel stronghold near the Lebanese border. • <i>September 2014</i>: US and five Arab countries launch air strikes against Islamic State around Aleppo and Raqqa. 	<ul style="list-style-type: none"> • <i>November 21, 2013</i>: President Yanukovich’s cabinet abandons an agreement on closer trade ties with EU, instead seeking closer co-operation with Russia. Small protests start. • <i>Early December 2013</i>: Protesters occupy Kiev city hall and Independence Square in dramatic style. 800,000 people rally in Kiev. • <i>December 17, 2013</i>: Vladimir Putin throws President Yanukovich an economic lifeline, agreeing to buy \$15bn of Ukrainian debt and reduce the price of Russian gas supplies by about a third. • <i>February 20, 2014</i>: Kiev sees its worst day of violence for almost 70 years. At least 88 people are killed in 48 hours. Video shows uniformed snipers firing at protesters holding makeshift shields. • <i>February 21, 2014</i>: President Yanukovich signs compromise deal with opposition leaders. • <i>February 22, 2014</i>: President Yanukovich disappears. Protesters take control of presidential administration buildings. Parliament votes to remove president from power and sets date for elections. Mr Yanukovich appears on TV to denounce “coup.” His arch-rival Yulia Tymoshenko is freed from jail. • <i>March 1, 2014</i>: Russia’s parliament approves President Vladimir Putin’s request to use force in Ukraine to protect Russian interests. • <i>March 16, 2014</i>: Crimea’s secession referendum on joining Russia is backed by 97% of voters. EU and US impose travel bans and asset freezes on officials from Russia and Ukraine in response. • <i>March 18, 2014</i>: President Putin signs a bill to absorb Crimea into the Russian Federation. • <i>March 28, 2014</i>: US President Barack Obama urges Moscow to “move back its troops” and lower tensions. • <i>July 30, 2014</i>: The EU and US announce new sanctions against Russia. • <i>September/October, 2014</i>: NATO reports a “significant” withdrawal of Russian troops from eastern Ukraine. Putin orders thousands of troops stationed near the Ukrainian border to return to their bases. Russia agrees to resume gas supplies to Ukraine over the winter in a deal brokered by the EU. • <i>February 15, 2015</i>: Ceasefire goes into effect (but violations quickly follow).

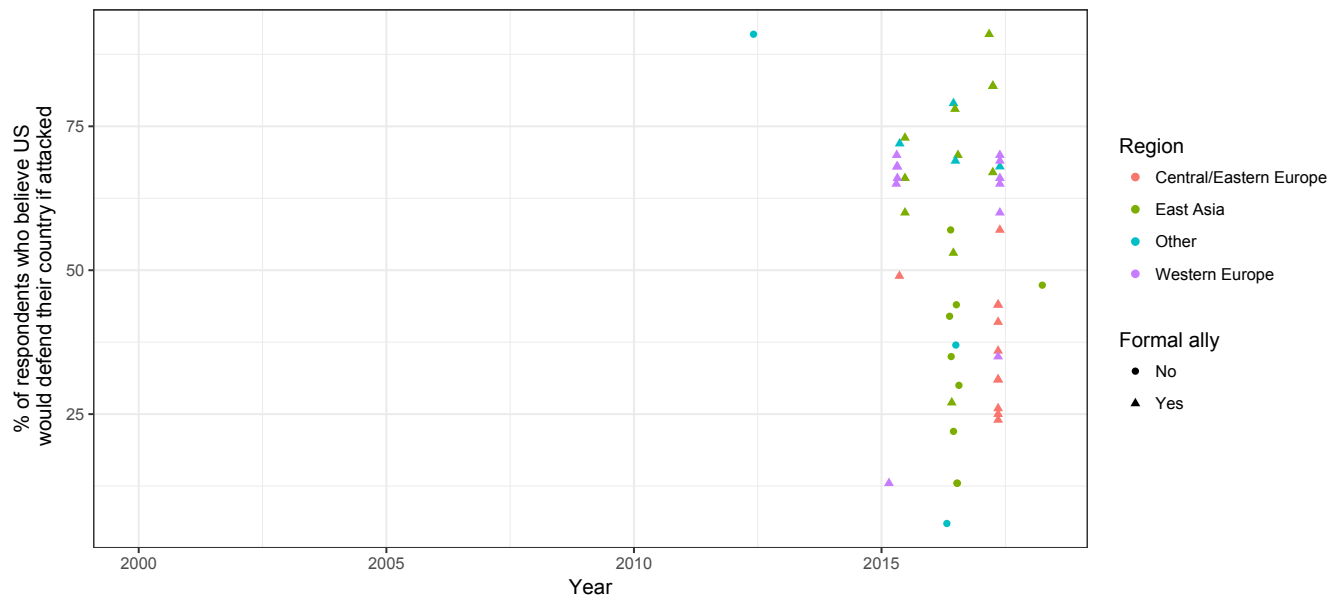
Figure 3: Crisis Timelines

2.1 Using Surveys to Study Reputation Costs

We fielded four surveys on three different populations: two samples of elites, and one mass public (See Figure 5 and Appendix §2).

Surveys are especially valuable for three reasons. First, if reputations are beliefs about the traits of particular actors, survey methods are particularly helpful at measuring the content of beliefs (Brutger and Kertzer, 2018, 4). They are better suited in this regard than large-N methods, which must rely on rough behavioral proxies in lieu of capturing the content of the beliefs themselves. Second, unlike much existing research, our sample here includes foreign leaders: current and former members of the Israeli Knesset. If the central focus of reputation-based arguments in US foreign policy is the reputational costs accrued *in the eyes of allies and adversaries*, the tools we employ here enable us to study this question directly. This is a notable departure from most work in IR, which tends to focus on the perspective of the *domestic* (public) audience, rather than the international one, even though the formal model whose microfoundations the experiments are testing is dyadic and even though the judgments of leaders and the public alike are implicated in our theories. Indeed, Jervis (1970, 8) long warned that we shouldn't assume that policymakers are concerned about reputation just for domestic political reasons. Understanding the *international* dimensions of reputation costs—among both leaders and the general public—using this method is thus particularly valuable.

Figure 4: Diagnostic polling questions, 2000-2015



Finally, although surveys are well-suited to studying the perceptual component of reputation in the *evaluation* phase, there are few existing surveys addressing the reputational questions we are interested in. Figure 4 plots reputation and credibility related questions from Pew, Gallup and several local polling from 2000-2015. We found neither questions that explicitly ask respondents about the link between an actor’s past actions and future behavior nor questions about the evaluation phase, which would have involved asking respondents to assess how the reputation of various actors have been affected by particular events. Instead, questions focused on future behavior—e.g., whether respondents believe the U.S. would defend their country—and even those did not appear with any regularity until 2015—*after* the crises in Crimea and Syria. As Krebs (2015, 136) notes, the questions polling firms choose to ask—and when—can itself be informative. That no polling firm in the past two decades thought to ask allied publics whether they believed the U.S. would defend them until *after* the crises in Crimea and Syria suggests the matter was not seen as being in question until then, underscoring the value of these two contemporary cases for the study of reputation.

Of course, surveys have tradeoffs on other dimensions. For example, surveys are not able to isolate causal factors in the same way as experimental methods. However, two factors make surveys the most appropriate choice in this particular case. First, our specific questions are not necessarily ones where experimental methods have an obvious advantage; for example, if our interest is in whether reputation costs change over time, even a well-designed experiment would be hard-pressed to causally identify the effect of “time.” Second, and more generally, any healthy research program ought to avoid relying too much on one particular method. In this case, surveys provide a critical balance to experimental methods by focusing on real-world events and helping to calibrate the size of effects observed in the “real world”.

2.2 Our Elite and Public Samples

“Elite” as it is used to describe survey/experimental samples in IR typically refers to either subject-matter experts (as in Tetlock, 2005, or TRIP surveys on IR scholars such as Milner and Tingley, 2013) or government officials or decision-makers (e.g., Hafner-Burton et al., 2014). We field our surveys on both, but with a crossnational twist. And while the benefits of elite samples are not always clear (Kertzer 2022 shows that “elite effects” often boil down to simple demographics), we employ them here in line with the recommendation of Kertzer and Renshon (2022, 538, 544) that “subjects correspond to the actors in the theory” and “directly implicate elites’ domain-specific

	Fielded:	Subject pool	location	N
Foreign leaders	July–October, 2015	Members of the Israeli Knesset	Israel	89
Foreign public I	October 6-9, 2015	Representative sample of Israeli public	Israel	1599
Foreign public II	January 18-25, 2016	Representative sample of Israeli public	Israel	1111
American IR scholars	September 28-30, 2015	TRIP survey	United States	692

Figure 5: Data Sources

expertise and experience.”

Our first elite sample consists of foreign decision-makers: current and former members of the Israeli Knesset (described in Figure 6; sampling procedure and representativeness analysis in Appendix §2). Israel is an important case in which to study perceptions of the United States’ reputation: despite the absence of a formal alliance, the US and Israel have a remarkably close informal alliance, reflected not just in the outsized volume of military aid the former gives to the latter, but in the extent to which both proponents and critics of US foreign policy in the Middle East see the two countries as bound together by a “special relationship” defined by shared perceptions of geopolitical threats in the region (Mearsheimer and Walt, 2007). Israeli decision-makers thus represent a more militarily powerful variant of a particular class of foreign leaders, occupying a similar structural position as Japan and South Korea in East Asia, Poland and Estonia in Eastern Europe, and so on: smaller powers dependent upon American backing in order to provide for their security, and thus the very type of ally that deterrence theory tells us should be invested in the maintenance of American credibility.

Our second elite sample is composed of subject matter experts: American IR scholars, recruited by the Teaching, Research and International Policy (TRIP) survey. This sample is “elite” in the sense of the term employed in previous elite surveys by Tetlock: respondents who—through training and/or experience—are subject-matter experts (a “cognitive model” of eliteness based on domain-specific expertise and knowledge, Kertzer and Renshon 2022). Of the 4,086 scholars polled, 694 responded, for a response rate of approximately 17%. Studying the reputational beliefs of American IR scholars is useful because they are knowledgeable about the events in Syria and Crimea and differences between them, and are familiar with the debate about the importance of reputation. The TRIP sample thus provides us with a valuable reference point with which to compare our responses from the Knesset members, to assess how similarly or differently reputation costs are assessed at home versus abroad, and by two samples that are “elite” in different ways: one through virtue of

Proportion of respondents		
Knesset Member:		
	Current	25%
	Former	75%
Exp. on Foreign Affairs/Defense Committee ...		
	... as backup or full member	67%
	... as full member	54%
Highest level of experience:		
	... not a Minister	58%
	... Deputy Minister	29%
	... Cabinet Member or higher	12%
Male		
		84%
Served in military		
		95%
Active combat experience		
		64%
	Mean	SD
Age	61	10.7
Terms in Knesset	3.0	2.1
Military Assertiveness	0.61	0.20
Right Wing Ideology	0.45	0.24
Hawkishness (Arab-Israeli conflict)	0.39	0.25
International Trust	0.40	0.26

Note: individual differences in bottom four rows scaled from 0-1.

Figure 6: Knesset Sample ($N = 89$)

expertise and training, the other through their experience, official responsibilities.

Our final source comes from two samples of the Israeli public (ISRAEL PUBLIC I & II), which were obtained using iPanel, an Israeli polling firm that has been used effectively by other recent surveys and experiments (e.g., [Kertzer, Renshon and Yarhi-Milo, 2021](#)). Both samples were representative of the Israeli Jewish population, and stratified based upon gender, age, living area and education. Fielding these two surveys on national samples (the first of which was fielded at roughly the same time as the TRIP survey and Knesset survey) is helpful for two purposes. First, because elite samples are inevitably small in size (by definition, the bigger the sample, the less elite it is), it allows us to replicate our findings on a much larger sample. Second, because one of the critiques of survey or experimental work is that the results are time-bound, replicating the study four months later allows us to test how reputational effects decay over time.

2.3 Survey Questions

In each study we presented respondents with eight items (ten, in the TRIP survey), spread across two grids, one in regard to Syria and the other to Crimea. In each grid, we focus specifically on the evaluation phase of reputation inferences. The first grid asks respondents to assess the reputation costs in Syria incurred by a variety of actors using a measure of reputation costs developed by Brutger and Kertzer:

Thinking about recent events in Syria, do you think the reputation of each of the following has been strengthened, weakened, or not affected by the way they have handled the situation in Syria?

The response options ranged from “Significantly weakened” (1) to “Significantly strengthened” (5), and respondents made reputational assessments for each of the following actors:

- The United States
- President Obama
- Russia
- The United Nations
- President Putin (TRIP survey only)

They then were presented with the same set of questions about events in Crimea.² Three points are of note here. First, to maximize breadth, we measure reputation costs with respect to a range of actors. Second, rather than priming participants to weight particular types of reputational considerations more than others — which would likely bias the results in favor of canonical models of reputation — we ask respondents to make general reputational assessments. This method dovetails with questions asked by major survey organization such as Pew, who often elicit beliefs about overall reputation rather than reputations for specific traits or qualities³ and aligns with the advice of Croco, Hanmer and McDonald (2021) to focus on more general quantities of interest in order to avoid overestimating effects. Finally, we do *not* prime participants by mentioning the infamous “red line” threat of 2012; we merely ask about the “recent events” in Syria. Doing so avoids putting our thumbs on the scale and biases us against finding any difference between the two crises.

3 Reputation Costs in Syria and Crimea

3.1 The Existence, Persistence and Magnitude of Reputation Costs

We begin with Figure 7, which plots the distribution of estimated reputational consequences incurred as a result of events in Crimea (top row) and Syria (bottom row) for our three actors: President Obama (left-hand column), the United States (right-hand column) and Russia (middle column).⁴

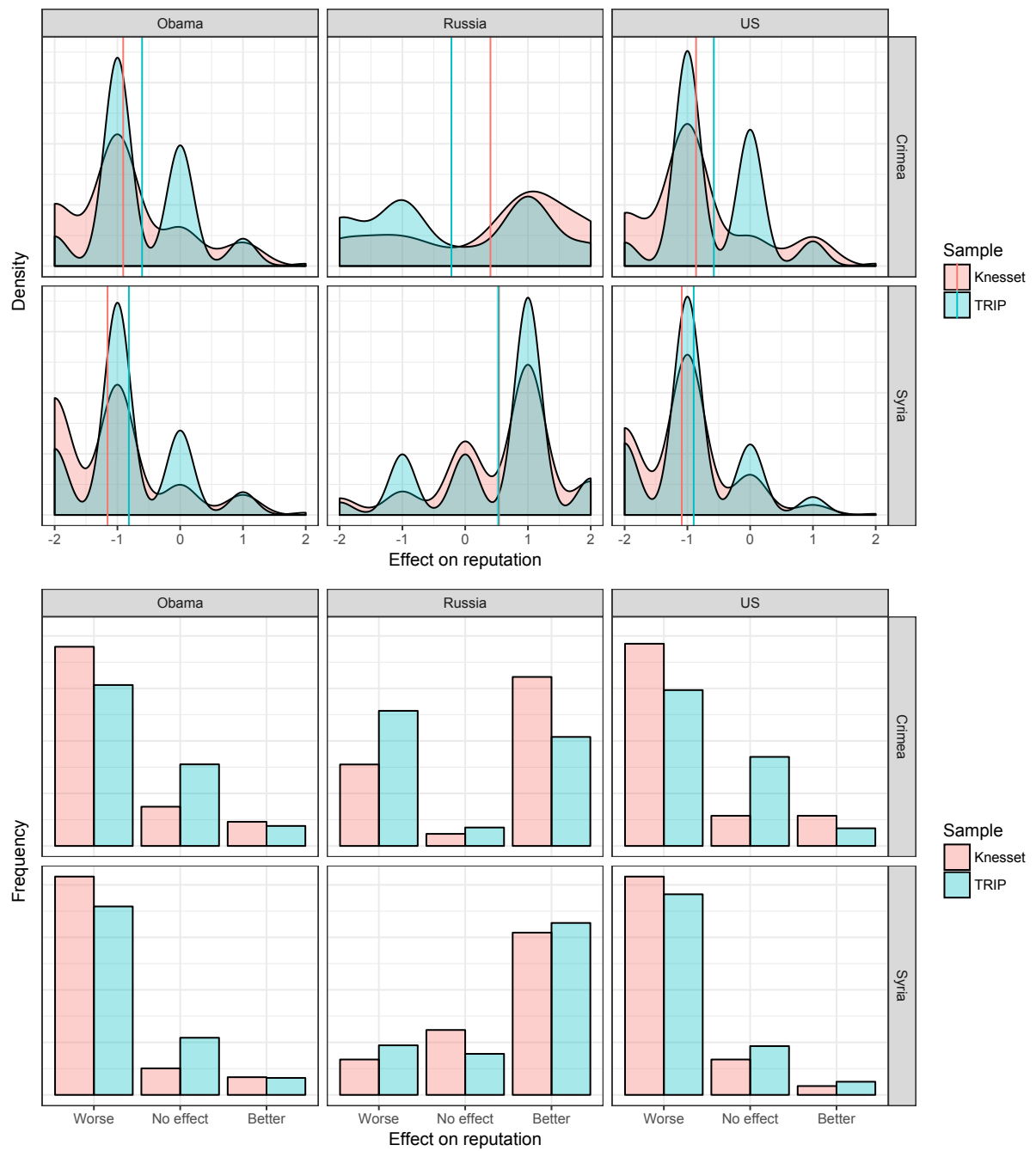
Broad Support for Existence of Reputation Costs: Figure 7 suggests strong support—across leaders and the public—for the argument that actors incur reputational costs from their actions on the international stage (H1). Our respondents perceived significant reputation costs incurred by the United States and its president and our two Israeli public samples agree (see Figure 8). Among elites, both IR scholars and Israeli leaders agree that Russia’s reputation was improved by events in Syria (though they disagree about the consequences of Crimea, which we discuss later). Importantly, this pattern is different from what is often reported in the literature. Reputation cynics like Press and Tang—and scholars writing more generally about reputation, like Yarhi-Milo—have consistently found evidence that leaders are concerned about their *own* reputations; we show here that *foreign*

²In the TRIP survey, participants were randomly assigned to receive only one of the two grids because of space constraints.

³See, for example <http://www.people-press.org/question-search/?qid=1827792&pid=51&ccid=51>.

⁴Data is re-centered so that positive values indicate reputational gains, and negative values indicate reputational costs. Complete results are presented in Appendix §1.

Figure 7: Comparing reputation costs in Syria and Crimea



The figure depicts the distribution of reputational costs incurred by President Obama, Russia, and the United States as a result of events in Crimea and Syria. In the bottom two panels, the x-axis has been rescaled to depict three types of respondents: those who perceived reputation costs, those who perceived reputation gains, and those who perceived no reputational change.

observers perceive reputation costs too.

Reputation Costs Are Stable Over Time: Our unique research design—we asked identical questions of both samples of the Israeli public, fielded three and a half months apart—permits us to evaluate H2 on the stability of reputational costs. During that time, a litany of events occurred that might have shifted the public’s beliefs.⁵ Despite these large and small events, our data reveals that—consistent with H2—the Israeli public’s assessment of reputation costs were highly stable between October 2015 and January 2016.⁶ The one exception to this is that perceptions of how much the reputation of the United States was harmed in the context of Syria did exhibit a statistically significant increase ($p < 0.009$), but the substantive effect (0.09, equivalent to 1.1 percentage points) was very small.

Reputation Costs Are Higher in Cases of Explicit Threats: To investigate H3—on the magnitude of reputation costs when red line threats are implicated—we compare the reputational consequences of Crimea (the top row) and Syria (bottom row). These crises occurred in the same geopolitical climate, at roughly the same time, involved many of the same actors and in both cases the U.S. did not use force. One significant difference, however, was Obama’s infamous explicit public threat, which is often understood as an indictment of audience cost theory, since the American public largely approved of the president’s decision to stay out (Gallup noted that American support for US intervention in Syria was “on track to be among the lowest for any intervention Gallup has asked about in the last 20 years.”⁷

Figure 7 shows that both samples see the US as paying steeper reputation costs in Syria than in Crimea, and our two Israeli public samples (Figure 8) reach a similar conclusion. This is consistent with Fearon (1994)’s canonical prediction that leaders should pay costs for making explicit coercive threats and not following through, yet it also encourages us to look beyond the narrow purview of a domestic punishment mechanism. As Gelpi and Grieco (2015) note, audience cost theory has been overly narrow in its focus on domestic electoral politics — such that failing to find evidence of leaders being ousted from office after making public threats is erroneously held as an indication that foreign policy misadventures are costless. Our results raise the specter of an alternative mechanism, in which leaders who back down on public threats suffer reputation costs in the eyes of *foreign*

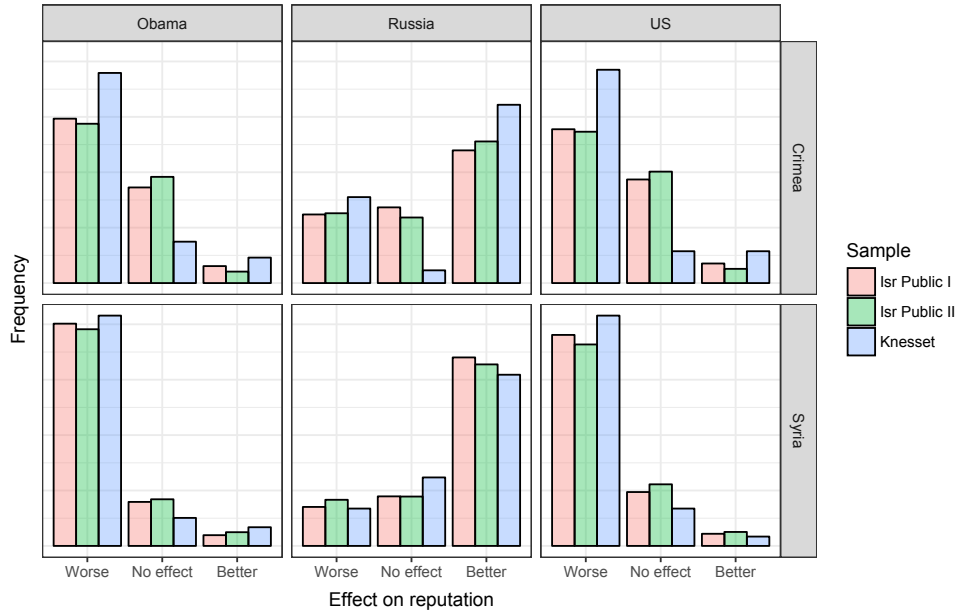
⁵For a more detailed timeline, see Figure 3.

⁶We test this formally by estimating a series of Wilcoxon rank-sum tests (see Appendix §1 for results).

⁷<http://news.gallup.com/poll/164282/support-syria-action-lower-past-conflicts.aspx>

audiences.

Figure 8: Comparing the Knesset and the two waves of Israeli public data



Of course, eliciting reactions to real-life paired crises such as Syria and Crimea sacrifices experimental control and limits our ability to make causal claims: Syria and Crimea are dissimilar in many ways besides the fact of whether a U.S. President made a threat: e.g., the cost of US intervention was lower in Syria than in Crimea. Moreover, geographic proximity might shape peoples’ reactions to the crises such that Israeli respondents may assess steeper reputation costs to the US in Syria because it is a neighboring country. However, that we see this same pattern among American IR scholars suggests at least that geographic proximity is not “doing the causal work” that we are assigning to failing to follow through on a threat.

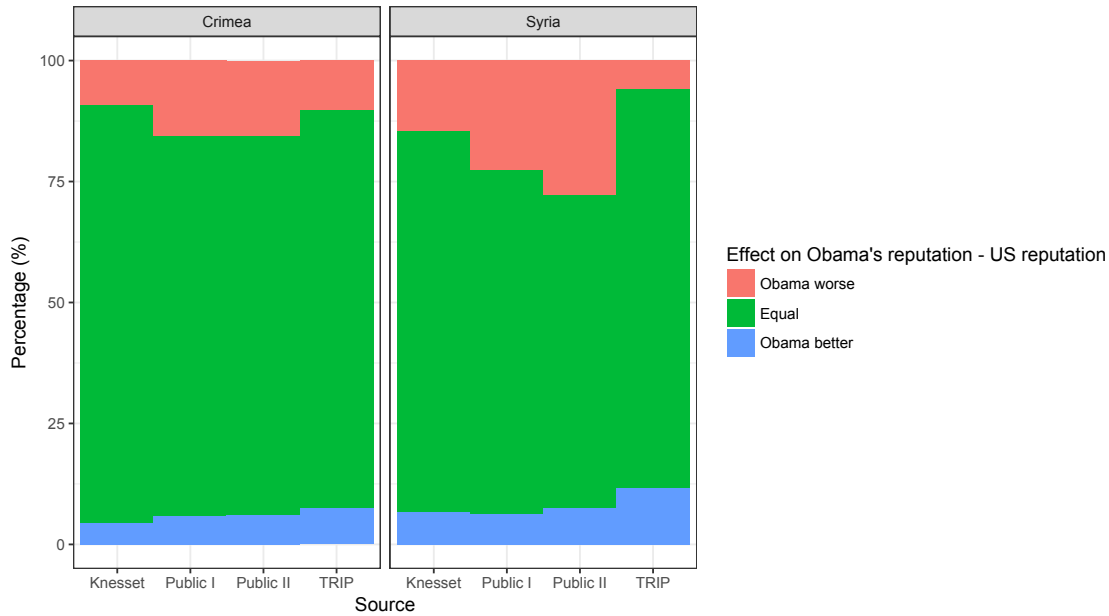
3.2 Variation in Reputation Costs Across Targets and Audiences

Next, we evaluate H4, which addresses variation in the scale of reputation costs attributed to different targets (actors). Put simply: to whom do these reputation costs adhere? Figure 9 compares the reputation costs attributed to President Obama versus the United States across the samples of Israeli leaders, US elites and the Israeli public. To facilitate a clearer interpretation, we calculate this as a within-subject estimate and trichotomize it to distinguish between those who allocate the same amount of reputational costs to both Obama and the United States (green), those who perceive

Obama as bearing less reputational cost than the United States (blue), and those who perceive Obama as bearing more reputational cost than the United States (red).

Reputation Costs Paid by Leaders and by States: Reputational judgments about the US and President Obama are highly correlated with one another: across all four samples, most respondents (65%- 86%, depending on the sample and crisis) attribute identical reputational costs to the president and the country as a whole. This is consistent with [Brutger and Kertzer \(2018\)](#), who find a similarly strong correlation between leader and country-specific assessments of reputation costs (in a domestic political context). Interestingly, however, we also see a sizable minority of respondents (14%-35%, depending on the sample and crisis) who allocate reputational costs differentially between the leader and country as a whole. Of the respondents who differentiated between country and leader-specific reputations, most tended to allocate greater costs to the leader than the country (especially in the Syrian context), with the exception being the TRIP survey, where American IR scholars were significantly more likely to give Obama a break compared to foreign respondents.

Figure 9: Calculating leader-specific reputation costs



Two points are noteworthy here. First, the 6-28% of respondents who allocate greater reputational costs to the leader are doing so in a manner consistent with extant theories of leader-specific reputations which posit that reputations are more likely to adhere to actors who are influential in the

Context	Actor	TRIP	Knesset	Difference	Wilcoxon test
Syria	US	-0.90	-1.09	0.19	$p < 0.03$
Syria	Obama	-0.82	-1.16	0.34	$p < 0.00$
Syria	Russia	0.53	0.54	-0.01	$p < 0.97$
Crimea	US	-0.58	-0.86	0.28	$p < 0.00$
Crimea	Obama	-0.53	-0.91	0.38	$p < 0.00$
Crimea	Russia	-0.22	0.40	-0.62	$p < 0.00$

Note: positive differences indicate foreign decision-makers perceive larger reputation costs than American IR scholars.

Figure 10: A significant “home-away” gap

state’s decision-making process; given the centrality of Obama’s involvement in the Syrian episode, it is not surprising that the reputational consequences should fall on his shoulders (Renshon, Dafoe and Huth, 2018). Our existing theories, however, have more difficulty explaining the existence of the 6-12% of respondents (depending on the sample and crisis) who give Obama a break and attribute greater reputational costs to the country as a whole, even though the episodes occurred during his presidency.⁸

Reputation Costs Vary Across Audiences: We now turn to assess if and how the magnitude of the reputation costs varies across audiences (H5), disaggregating along two dimensions: foreign-domestic audiences and elite-public audiences. As Figure 7 indicates, there is a strong support for H5: American IR scholars perceive significantly less severe costs to America’s reputation from Syria and Crimea than our sample of foreign decision-makers do (Figure 10). The same is also true when we compare American IR scholars with the two Israeli public samples.

Why do we American IR scholars go easier on the United States when assessing reputation costs? One possible explanation is motivated reasoning or ethnocentrism (Herrmann, 2017), though if elites tend to be more cosmopolitan and less nationalist than the public, any explanation based on the nationalist tendencies of American IR scholars would be correspondingly less likely. Another interpretation of this pattern is attribution bias: Americans might perceive their country’s reputation costs as being smaller than foreign observers do because they place a greater weight on environmental constraints when it comes to their own country’s behavior (Mercer, 1996). Other explanations are plausible, but do not fit our empirical patterns. For example, though IR scholars might be more

⁸This is not simply a result of liberals giving Obama a break: among TRIP sample, political ideology is not a significant predictor for either Crimea or Syria; among two Israeli public samples, results vary somewhat by sample and context, but politically knowledgeable respondents, older respondents, and non-Ashkenazi respondents are generally less likely to give Obama a break.

likely to self-define as “realists” focused on capabilities and interests, the realists in our sample never estimate significantly smaller reputation costs than non-realist respondents.⁹ Similarly, given the ideological leanings of the Academy, American IR scholars might be more likely to support a Democratic administration (Rathbun, 2012). However, while right-leaning American IR scholars perceive significantly larger reputation costs for the United States than left-leaning ones, we show in Appendix §1.1 that our “home-away gap” cannot be due entirely to the ideological leanings of the Academy. Finally, it might be that our results reflect a predisposition of Israelis to care more about reputation compared to Americans, though recent work suggests no significant difference between the two publics in their willingness to “fight for reputation” (Yarhi-Milo, 2018).

Finally, our design allows us to assess variation in reputation costs across samples of elites and the general public through comparison of identical survey questions fielded to members of the Israeli Knesset and a (much larger) representative sample of the Israeli public simultaneously. This addresses both a theoretical question about reputation costs, as well as longstanding empirical concerns about the nature of survey samples, which tend to coalesce around the question: would our results look different if we had gained access to different (and in IR, more elite) subjects?

Figure 8 provides an overview of the distributions for each response item and Figure 11 provides more formal tests of congruence. There we show that, for both public samples, we fail to detect statistically significant differences between elites and the public on Syria. For Crimea, we find that the Israeli public perceives slightly less steep reputation costs for the US than the Knesset members do, although not for Obama, where elites and publics do not significantly differ from one another. On the whole, then, foreign elites and their publics seem to assess reputation costs rather similarly: both groups saw Syria (and the failure to follow through on a threat) as hurting Obama’s reputation as well as that of the United States and the events in Crimea as harming the reputation of the U.S. and Obama, with only slight differences in the magnitude for the country’s reputation.

Conclusion

In this research note, we have sought to make two contributions. First, we suggest that the “reputation debate” in IR is really two distinct debates, with some scholars investigating the *evaluation* phase (do observers draw reputational inferences in response to past actions in the manner our

⁹In fact, realists perceive *larger* reputation costs than non-realists in the case of Obama’s reputation costs for Crimea.

Context	Actor				Wilcoxon test I	Wilcoxon test II
		Knesset	Public I	Public II	(Knesset/Public I)	(Knesset/Public II)
Syria	US	-1.09	-1.03	-0.94	$p < 0.65$	$p < 0.15$
Syria	Obama	-1.16	-1.20	-1.15	$p < 0.66$	$p < 0.99$
Crimea	US	-0.86	-0.70	-0.71	$p < 0.04$	$p < 0.04$
Crimea	Obama	-0.91	-0.80	-0.81	$p < 0.21$	$p < 0.20$

Figure 11: Elite-public congruence: numbers under Knesset/Public columns indicate how much reputation of actor was hurt in a given context.

theories suggest?) and others focusing on the *diagnostic* phase (how much do observers weight past actions when formulating present behavior?). Second, since reputations are beliefs, we introduce a new type of evidence to the debate, leveraging a unique survey research design asking identical questions from respondents from four samples in two countries. Together, they shed insight on two sets of critical questions relating to the evaluation phase of reputation: (1) do reputational beliefs form (and if so, how do they shift over time and in response to different types of events)? (2) how do reputation costs vary across both targets of inference (different actors) and across audiences?

A number of important findings emerge from our data. First, all of our respondents see the United States and President Obama as having incurred significant reputational costs as a result of both Syria and Crimea. Not all events affect reputations equally, however. Every one of our samples saw the U.S. as paying reputation costs in Syria, where—in a rare outlier event—a leader made a clear, dramatic “red line” threat and did not follow through. These costs were steeper compared to Crimea, underscoring the importance observers attribute to commitments that explicitly stake the leader’s reputation for resolve.

We also find that reputation costs are in the eye of the beholder, with our results suggesting a “home-away gap” in which U.S. foreign policy scholars were far more sanguine about reputation costs incurred by their own country than foreign observers. While our research design cannot adjudicate between explanations for this pattern—whether attribution bias, a more liberal sample, or simple ethnocentrism—the implications do help us refine our understanding of reputation. Critics of Obama’s foreign policy legacy (particularly on the political right) suggested that the administration’s inaction in Crimea and Syria sullied the country’s reputation, yet Obama did not appear to pay any consequences domestically for his inaction (Dueck, 2015) leading some IR scholars to suggest a potential weakness in audience cost theory, which is sometimes criticized for exaggerating the costliness of empty threats (Snyder and Borghard, 2011). However, if audience cost models in IR

focus our attention on the *domestic* consequences of failing to follow through on threats, our data suggest a renewed focus on the international side of the equation that is more in line with the writing of Schelling and Jervis on this topic (and recent work by [Zhang 2018](#)).

Third, our surveys helped refine our understanding of how reputations “attach” to different actors. While first-generation work assumed that countries were the sole owners of their reputations, more recent work has shown that reputations may adhere to leaders as well. While the vast majority of our respondents attributed responsibility about equally between Obama and the United States, a significant minority “blamed” Obama more than the United States, a finding that fits neatly with [Renshon, Dafoe and Huth \(2018\)](#), who argue that foreign policy is exactly where we would expect observers to attribute more responsibility to individual leaders.

References

- Brutger, Ryan and Joshua D. Kertzer. 2018. "A Dispositional Theory of Reputation Costs." *International Organization* 72(3):693–724.
- Brutger, Ryan, Joshua D Kertzer, Jonathan Renshon, Dustin Tingley and Chagai M Weiss. 2022. "Abstraction and detail in experimental design." *American Journal of Political Science* .
- Casler, Don and Richard Clark. 2021. "Trade Rage: Audience Costs and International Trade." *Journal of Conflict Resolution* 65(6):1098–1130.
- Clare, Joe and Vesna Danilovic. 2010. "Multiple Audiences and Reputation Building in International Conflicts." *Journal of Conflict Resolution* 54(6):860–882.
- Crescenzi, Mark J.C. 2018. *Of Friends and Foes: Reputation and Learning in World Politics*. New York, NY: Oxford University Press.
- Croco, Sarah E, Michael J Hanmer and Jared A McDonald. 2021. "At what cost? Reexamining audience costs in realistic settings." *The Journal of Politics* 83(1):8–22.
- Dafoe, Allan, Jonathan Renshon and Paul Huth. 2014. "Reputation and Status as Motives for War." *Annual Review of Political Science* 17:371–393.
- Dueck, Colin. 2015. *The Obama doctrine: American grand strategy today*. New York, NY: Oxford University Press.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88(3):577–592.
- Fisher, Max. 2017. "Do U.S. Strikes Send a 'Message' to Rivals? There's No Evidence." New York Times.
- Gelpi, Christopher and Joseph M. Grieco. 2015. "Competency Costs in Foreign Affairs: Presidential Performance in International Conflicts and Domestic Legislative Success, 1953–2001." *American Journal of Political Science* 59(2):440–456.
- Goldberg, Jeffrey. 2016. "The Obama doctrine."
URL: <https://www.theatlantic.com/magazine/archive/2016/04/the-obama-doctrine/471525/>
- Hafner-Burton, Emilie M., Brad L. LeVeck, David G. Victor and James H. Fowler. 2014. "Decision Maker Preferences for International Legal Cooperation." *International Organization* 68(4):845–876.
- Harvey, Frank P. and John Mitton. 2017. *Fighting for Credibility: US Reputation and International Politics*. University of Toronto Press.
- Herrmann, Richard K. 2017. "How Attachments to the Nation Shape Beliefs About the World: A Theory of Motivated Reasoning." *International Organization* 71(S1):S61–S84.
- Hopf, Ted. 1994. *Peripheral visions: Deterrence theory and American foreign policy in the third world, 1965-1990*. Ann Arbor, MI: University of Michigan Press.
- Jervis, Robert. 1970. *The Logic of Images in International Relations*. Princeton, NJ: Princeton University Press.
- Jervis, Robert, Keren Yarhi-Milo and Don Casler. 2021. "Redefining the Debate over Reputation and Credibility in International Security: Promises and Limits of New Scholarship." *World Politics* 73(1):167–203.

- Kertzer, Joshua D. 2022. "Re-assessing elite-public gaps in political behavior." *American Journal of Political Science* 66(3):539–553.
- Kertzer, Joshua D and Jonathan Renshon. 2022. "Experiments and surveys on political elites." *Annual Review of Political Science* 25:529–550.
- Kertzer, Joshua D., Jonathan Renshon and Keren Yarhi-Milo. 2021. "How Do Observers Assess Resolve?" *British Journal of Political Science* 51(1):308–330.
- Krebs, Ronald R. 2015. *Narrative and the Making of US National Security*. Cambridge, UK: Cambridge University Press.
- Lupton, Danielle L. 2018. "Reexamining Reputation for Resolve: Leaders, States, and the Onset of International Crises." *Journal of Global Security Studies* 3(2):198–216.
- Lupton, Danielle L. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca, NY: Cornell University Press.
- McManus, Roseanne W. 2018. "Making It Personal: The Role of Leader-Specific Signals in Extended Deterrence." *Journal of Politics* 80(3):982–995.
- Mearsheimer, John J. and Stephen M. Walt. 2007. *The Israel Lobby and U.S. Foreign Policy*. New York: Farrar, Straus and Giroux.
- Mercer, Jonathan. 1996. *Reputation and international politics*. Ithaca, NY: Cornell University Press.
- Milner, Helen V. and Dustin H. Tingley. 2013. "The Choice for Multilateralism: Foreign Aid and American Foreign Policy." *Review of International Organizations* 8(3):313–341.
- Press, Daryl G. 2005. *Calculating credibility: How leaders assess military threats*. Ithaca, NY: Cornell University Press.
- Rathbun, Brian C. 2012. "Politics and Paradigm Preferences: The Implicit Ideology of International Relations Scholars." *International Studies Quarterly* 56(3):607–622.
- Rathbun, Brian C. 2014. *Diplomacy's Value: Creating Security in 1920s Europe and the Contemporary Middle East*. Ithaca, NY: Cornell University Press.
- Renshon, Jonathan, Allan Dafoe and Paul K. Huth. 2018. "Leader Influence and Reputation Formation in International Politics." *American Journal of Political Science* 62(2):325–339.
- Renshon, Jonathan, Keren Yarhi-Milo and Joshua D. Kertzer. 2023. "Democratic Reputations in Crises and War." *Journal of Politics* 85(1):1–118.
- Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.
- Shannon, Vaughn P. and Michael Dennis. 2007. "Militant Islam and the Futile Fight for Reputation." *Security Studies* 16(2):287–317.
- Snyder, Jack and Erica D. Borghard. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105(3):437–456.
- Tang, Shiping. 2005. "Reputation, Cult of Reputation, and International Conflict." *Security Studies* 14(1):34–62.
- Tetlock, Philip E. 1999. "Theory-Driven Reasoning about Plausible Past and Probable Futures in World Politics: Are We Prisoners of Our Preconceptions?" *American Journal of Political Science* 43(2):335–366.

- Tetlock, Philip E. 2005. *Expert Political Judgment: How Good is It? How Can We Know?* Princeton, NJ: Princeton University Press.
- Tomz, Michael. 2007. "Domestic audience costs in international relations: An experimental approach." *International Organization* 61(4):821–840.
- Weisiger, Alex and Keren Yarhi-Milo. 2015. "Revisiting Reputation: How Past Actions Matter In International Politics." *International Organization* 69(2):473–495.
- Wu, Cathy Xuanxuan and Scott Wolford. 2018. "Leaders, states, and reputations." *Journal of Conflict Resolution* 62(10):2087–2117.
- Yarhi-Milo, Keren. 2018. *Who Fights for Reputation in International Politics? Leaders, Resolve, and the Use of Force*. Princeton, NJ: Princeton University Press.
- Zhang, Ketian. 2018. "Cautious Bully: Reputation, Resolve, and Beijing's Use of Coercion in the South China Sea." *International Security* 44(1):117–159.

Reputations in the Wild

Supporting Materials

Contents

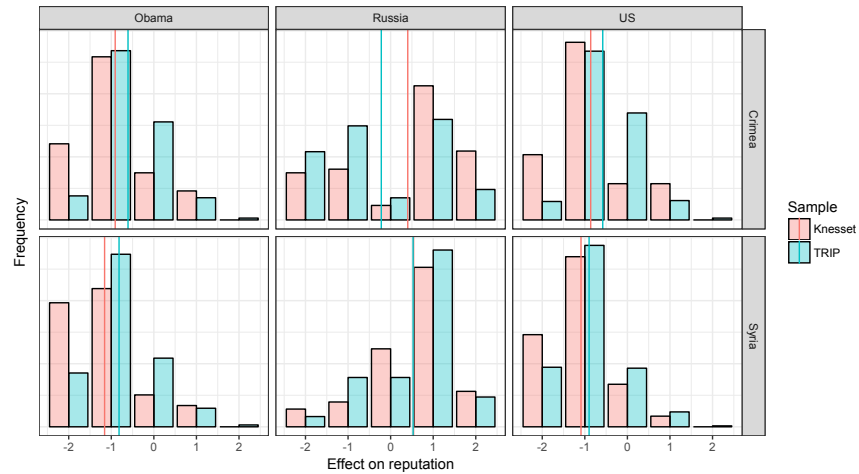
1	Supplementary Analysis	1
	Table 1: Reputation costs stable over time	1
	Figure 1: Comparing reputation costs in Syria and Crimea	2
	Figure 2: Comparing the Knesset and the two waves of Israeli public data	2
	Figure 3: Raw distribution of reputation beliefs: Knesset	3
	Figure 4: Raw distribution of reputation beliefs: Israeli public I	3
	Figure 5: Raw distribution of reputation beliefs: Israeli public II	3
	Figure 6: Raw distribution of reputation beliefs: American IR scholars	4
	Figure 7: Joint distribution of leader-specific reputation costs: American IR scholars	4
1.1	Ideological decomposition of “home-away” gap	5
	Figure 8: The “home-away” gap persists even when accounting for ideological distance	6
2	Samples	6
2.1	Knesset - Overview	6
2.2	Knesset Studies	7
	Figure 9: Recruitment Letter	8
	2.2.1 Participant verification protocol	8
	2.2.2 Sample characteristics and representativeness	9
	Table 2: Knesset Sample Characteristics	9
3	Content Analysis	9
	Table 3: Sample representativeness tests	10

1 Supplementary Analysis

Table 1: Reputation costs stable over time

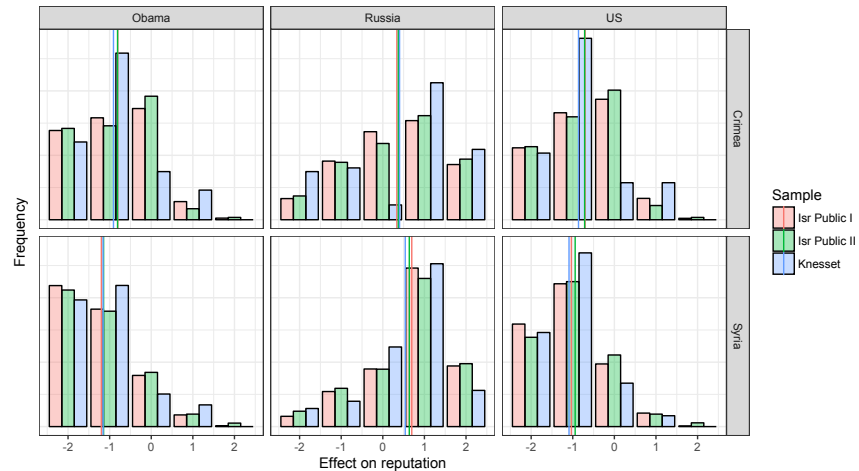
Context	Actor	Public II - Public I	Wilcoxon test
Syria	US	0.09	$p < 0.009$
Syria	Obama	0.05	$p < 0.25$
Crimea	US	-0.01	$p < 0.85$
Crimea	Obama	-0.005	$p < 0.98$

Figure 1: Comparing reputation costs in Syria and Crimea



The figure depicts the distribution of reputational consequences incurred by a variety of actors (President Obama, Russia, and the United States) as a result of events in Crimea (the top row) and Syria (the bottom row). The figure replicates Figure 3 from the main paper, but without trichotomizing the dependent variable; negative values indicate reputation costs, and positive values reputation benefits. The estimates from foreign decision-makers — members of the Knesset — are depicted in peach, and from American IR scholars in turquoise. The two vertical lines in each panel denote the means of each distribution. Thus, the figure depicts at least two important dynamics: first, as comparing the top and bottom rows shows, both samples perceive the US and Obama as incurring steeper reputational costs in Syria (where the President issued an explicit public threat) than Crimea. Second, as comparing the peach and turquoise distributions in each panel shows, American IR scholars consistently perceive a lower magnitude of reputational costs when compared to foreign decision-makers.

Figure 2: Comparing the Knesset and the two waves of Israeli public data



The figure depicts the distribution of reputational consequences incurred by a variety of actors (President Obama, Russia, and the United States) as a result of events in Crimea (the top row) and Syria (the bottom row). The figure replicates Figure 5 from the main paper, but without trichotomizing the dependent variable; negative values indicate reputation costs, and positive values reputation benefits. The estimates from the two waves of Israeli public samples are depicted in red and green, while the results from foreign decision-makers — members of the Knesset — are depicted in blue. The three vertical lines in each panel denote the means of each distribution. Thus, the figure depicts at least two important dynamics: first, as comparing the top and bottom rows shows, both samples perceive the US and Obama as incurring steeper reputational costs in Syria (where the President issued an explicit public threat) than Crimea. Second, as comparing the peach and turquoise distributions in each panel shows, American IR scholars consistently perceive a lower magnitude of reputational costs when compared to foreign decision-makers.

Figure 3: Raw distribution of reputation beliefs: Knesset

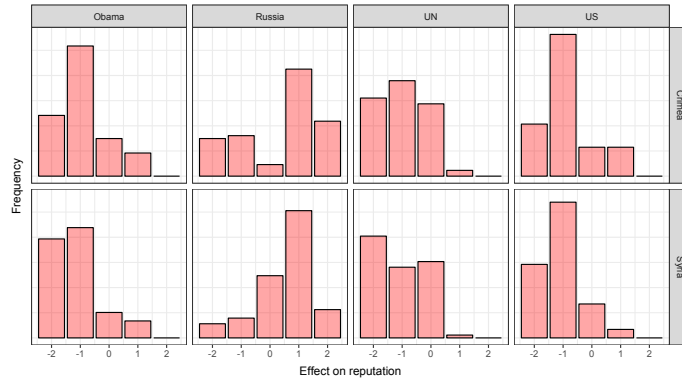


Figure 4: Raw distribution of reputation beliefs: Israeli public I

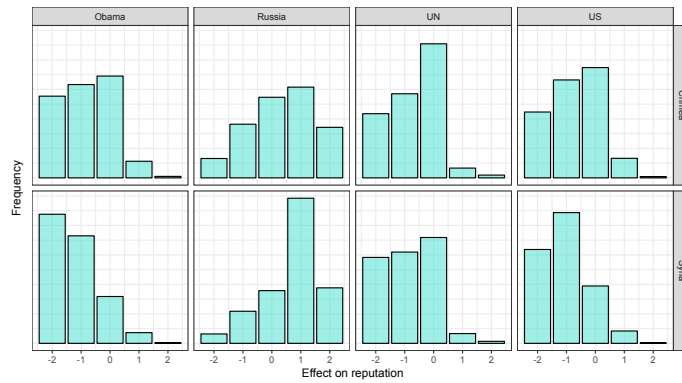


Figure 5: Raw distribution of reputation beliefs: Israeli public II

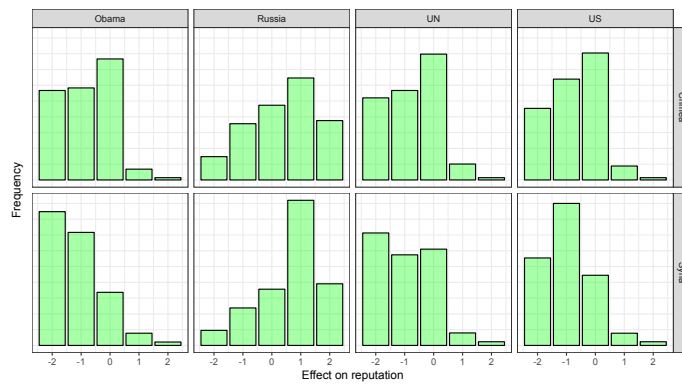


Figure 6: Raw distribution of reputation beliefs: American IR scholars

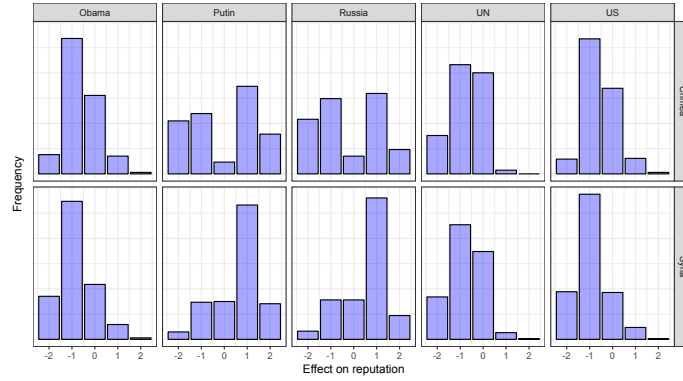
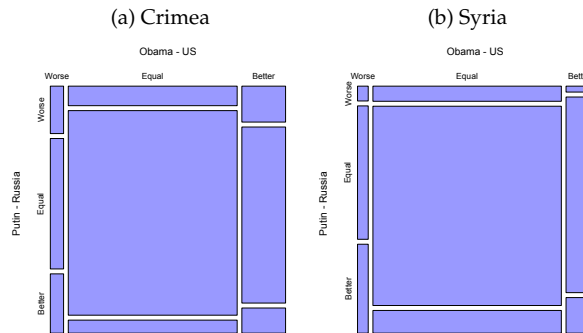


Figure 7: Joint distribution of leader-specific reputation costs: American IR scholars



Because the TRIP sample includes assessments of reputation costs for both the US and Obama, on the one hand, and Russia and Putin, on the other, we can analyze the joint distribution of leader-specific reputation costs for each leader within each conflict. The x-axis in each panel is comparable to the coding scheme in Figure 4 in the main text, calculating leader-specific reputation costs for Obama in which Obama's reputation costs are either worse, equal, or better than those allocated to the United States. The y-axis in each panel does the same, but comparing Putin's reputation costs to those allocated to Russia.

1.1 IDEOLOGICAL DECOMPOSITION OF “HOME-AWAY” GAP

One of the consistent patterns we find in our results in the main paper is the presence of a “home-away” gap between our domestic and foreign elite samples, with Americans consistently providing smaller assessments of the reputation costs than their foreign counterparts. There are a number of potential explanations we test for in the paper, ranging from paradigmatic attachments (perhaps American IR scholars are more likely to define as realists, who have tended to be critical of the American foreign policy establishment’s preoccupation with reputation) to attribution bias (perhaps foreign decision-makers discount situational factors at a greater rate when explaining American behavior, leading to larger assessments of reputation costs). One potential explanation involves an ideological gap between our two elite samples, with our Israeli sample more right-leaning than our sample of American IR scholars, who tend to be more liberal. The empirical strategy we employ to test this proposition involves measuring each sample’s political ideology, and using entropy balancing (Hainmueller, 2012) to reweight the American elite sample to look more like the Israeli elite sample in terms of political ideology, measuring the extent to which the magnitude of the home-away gap shrinks.

There are two methodological challenges here. The first is that whereas our Knesset sample measures ideology using the standard seven-point left-right scale used in survey research in political science (ours was based on that used by the Israeli National Election Study), the TRIP survey measures ideology using two separate five-point scales, one capturing ideology on social issues, and the other on economic issues, response options for each of which ranged from “Very liberal” to “Very conservative”.¹ The second is the question of measurement equivalence in cross-national survey research more generally (Davidov, 2009; Davidov et al., 2014). While the latter consideration is an indelible feature of cross-national work, such that the analysis below should be interpreted with caution, to address the former we use the empirical strategy described below.

1. Calculate a mean ideology score for each TRIP survey respondent, which is an average of their social and economic ideology.
2. Trichotomize the ideology score for the Knesset sample, as well as each of the three ideology scores for the TRIP sample (mean ideology, social ideology, and economic ideology) into liberal (all responses below the scale midpoint), moderate (all responses at the scale midpoint), and conservative (all responses above the scale midpoint).
3. Calculate the proportion of Knesset respondents who are liberal, moderate, and conservative, respectively.
4. Use entropy balancing to generate weights for the TRIP respondents, to reweight the TRIP sample to the target ideology proportions from the Knesset sample
5. Repeat the above step for each of the three ideology measures from TRIP, trimming weights to reduce the impact of extreme values.²
6. Calculate the average reputation cost in the TRIP sample, using each of the three weighting schemes, to ensure that the results are not the artifact of a particular ideology measure.

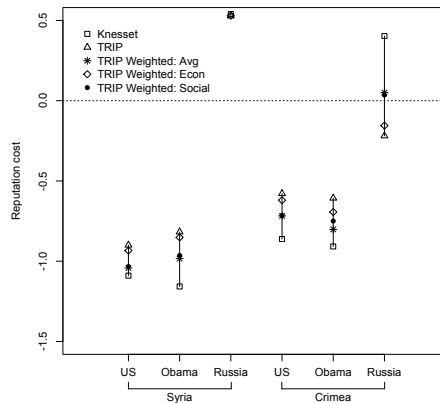
Figure 8 presents the results. Each of our six main measures of reputation costs is presented on the x axis, and the average reputation costs are depicted on the y-axis; more negative numbers indicate steeper assessments of reputation costs, while positive numbers indicate the actor’s reputation was perceived as improving. The results for each sample are depicted with a different symbol: the Knesset results with an open square, the unweighted results with the open triangle,

¹On debates about the dimensionality of political ideology, see Jost, Federico and Napier 2009; Duckitt and Sibley 2010; Feldman 2013

²See Huff and Kertzer (2018) for another application of this procedure.

and the three weighted TRIP results with an asterisk, diamond, and filled circle, based on the ideology measure being used. The vertical line connecting the open square and triangle depict the magnitude of the “home-away gap” reported in the main text. As noted in the main text, the US and President Obama are consistently attributed steeper reputation costs by our Kneset sample than our TRIP sample, both with respect to Syria and Crimea; Russia is also given strikingly more positive reputational assessments by Israeli leaders than by American IR scholars in the context of Crimea.

Figure 8: The “home-away” gap persists even when accounting for ideological distance



The more interesting question is the extent to which the gap shrinks once we reweight the American sample to ideologically look more like the Israelis. As noted above, we use three sets of weights: one based on economic ideology (depicted by the open diamond), one based on social ideology (depicted by the filled circle) and one based on the average score of the two (depicted by the asterisk). In general, the economic ideology reweighting consistently shows results close to the unweighted TRIP sample (the diamond is always close to the triangle), while the social ideology reweighting shows results that are more like the Kneset sample. In all cases, however (including with the average ideology score), a home-away gap persists between the samples, sometimes (as in the case of Russia) very strikingly so. These findings suggest that while ideological divides between a left-leaning sample of American IR scholars, and a more conservative sample of Israeli politicians, is likely responsible for some of the gap we perceive in reputation costs, it is unlikely to be accountable for all of it. As we note in the main text, there is another dynamic to consider here as well, which is that if we do think the home-away gap is partially a function of ideology, the fact that our Kneset sample includes former Kneset members in it (who on the whole are less conservative than current MKs are, as a result of Israel’s rightward shift), we might expect that our study is underestimating the magnitude of this gap, rather than overestimating it.

2 Samples

2.1 KNESSET - OVERVIEW

We began with a “universe” of Kneset members that included anybody that had served since 1996, including members of the 14th - 20th Knessets, a total of 408 individuals. After verifying the contact information we were left with 288 potential candidates (all 120 current members plus 168 former members). Of those 288 potential subjects, 89 participated, leaving us with a 31% response rate, relatively high for surveys of this type. The Kneset sample is described in Table ?? and our recruitment procedures are described in greater detail below. 25% of our participants were currently

serving in the Knesset; the rest of the sample (75%) was composed of former Knesset members. They were also highly experienced in IR-relevant contexts: 64% had active combat experience, and 67% had experience serving as members of the Knesset's Foreign Affairs and Defense committee. In addition to experience along the dimension of military strategy, they also had considerable political experience: our participants served an average of 3 terms in Parliament (with some as high as 9 terms). While 61% of the Knesset subjects had never served as a Minister, 29% had been at least a Deputy Minister, and fully 12% of our sample was in our highest category of elite experience, such that our participants include individuals who had served as Cabinet members, and even Prime Minister.³ Indeed, since our study was fielded, over a quarter of the sitting MKs who were backbenchers when they took our study have been promoted to Minister.

2.2 KNESSET STUDIES

The recruitment process began by compiling a dataset of all 415 individuals who had served as members of Israeli Parliament (i.e., the Knesset) from the beginning of the 14th Knesset in June 1996 through the 20th Knesset (the current Knesset) that was sworn in in March 2015. We compiled a data set that included the following information about our population: full name, party affiliation while in Knesset, names of all Knesset committees on which (s)he served, number of terms served, whether (s)he served as a minister in the government, and if so, what portfolios (s)he held, whether (s)he was a member of the Cabinet.

Contact information for our participants was obtained through a variety of channels, including the Secretary of the Knesset, the Knesset Channel, the different parties' leadership offices in the Knesset and other government offices where former Knesset members are currently employed. Email addresses for all current members of the Knesset were obtained through the Secretary of the Knesset. To verify whether the contact information we obtained was correct, we either called or emailed all the former Knesset members from the last twenty years and asked them if they would be interested in taking a "10 minute electronic survey by a team of professors from leading American Universities." 30.6% of the initial population was removed from the sampling frame at this stage, either because the members were deceased, were too sick to participate, or because their contact information was out of date and newer contact information could not be found. This process left us with a sample of 288 potential candidates to take our survey. This pool included all 120 current members of the Knesset along with 168 former members whose contact information was available.

On July 10, 2015, we executed a soft launch of our on-line survey. The survey included a recruitment email, written in Hebrew (reproduced below), a link to our on-line survey, and an individual six-digit password that was pre-assigned to each member. In the following days, we emailed the invitation to all current and former members in our dataset. A few weeks later, we sent a reminder email to those who had not responded to the survey. We sent a third round of reminders a few weeks later. In between these rounds, we phoned former and current Knesset members or their assistants to remind them to take the survey. In early August, the Director of Academic Affairs at the Knesset, together with the Secretary of the Knesset, sent an email to all current Knesset members encouraging them to take the survey, repeating essentially the same information we provided in the introductory email.

In addition to the on-line survey, we created identical hard-copy versions of our survey. In mid-August we sent those who had not responded to our survey a reminder email and attached an electronic copy of our survey that could be opened in Microsoft Word. Respondents were given the option of either faxing or emailing the completed survey back to us. That same six-digit code was the only identifying information on the paper copies of the survey, allowing us to track completion among our sample population. Members of our research team also traveled to the Knesset on four separate occasions to invite current members to participate.

³Below, we conduct supplementary analyses showing that our sample is fairly representative, although not surprisingly, current Knesset members were less likely to participate than former members.

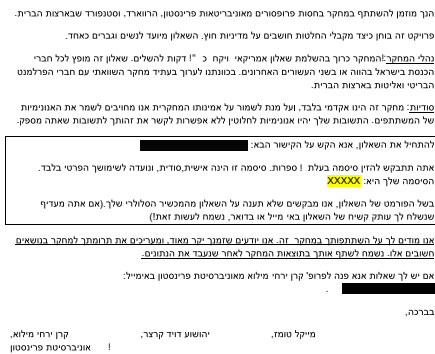


Figure 9: Recruitment Letter

The entire recruitment process was done in Hebrew. Two Hebrew-speaking research assistants and one member of the research team who is a native Hebrew speaker corresponded with the members of the Knesset or their assistants. Participants were informed that there would receive no financial reward for taking the survey, but that we would be happy to share with them the results of the survey. Moreover, participants were promised full anonymity: with the exception of the research team, participants were assured that identifiable information would not be released or reported.

2.2.1 Participant verification protocol

We took several steps to increase our confidence that the current and former decision-makers participated in the study rather than members of their staff. First, in the introductory email we explicitly indicated that the questionnaire should be fielded by the decision-maker himself, and not by members of his or her staff. We explained that the code we provided to access the on-line survey was personal, and should not be shared with others. Importantly, we did not offer any material incentives for filling out the survey, to dissuade decision-makers and assistants for taking the survey for those material reasons.

Second, in the survey itself we asked the participants to enter their complete date of birth. This allowed us to compare this information with the date of the decision-makers in official Knesset records. Third, for the 75% of our sample consisting of former Knesset members, a Hebrew-speaking research assistant and one of the authors were both in touch with the decision-maker directly via phone or email, and confirmed with him/her that they were the ones taking the survey. Anecdotally, our research team found that many of our participants were quite eager for the opportunity to opine on issues of foreign policy to an outside audience.

In the case of some current Knesset members, after receiving approval from their parliamentary assistant, a Hebrew-speaking research assistant from our team or one of the authors gave the Knesset members the survey directly and picked it up from them within a two-hour window. However, some Knesset members wished to maintain their anonymity and thus were not in direct contacted with the research team.

Finally, although we follow best practices, as is always the case with elite experiments, we should note that decision-makers who wished to “cheat” and delegate their participation to others could have probably found ways to do so. However, the combination of the types of questions asked in the survey, the absence of material compensation for survey completion, our explicit request the survey not be filled out by others, and the enthusiastic response to our survey from most of the decision-makers who took the survey leave us confident that the vast majority of them participated directly.

2.2.2 Sample characteristics and representativeness

Table 2 presents basic descriptive statistics for our Knesset sample. As the table shows, the sample is unusually “elite” by the standards of many experiments conducted in international relations: two-thirds of the sample has experience on the foreign affairs and defense committee, and over 40% served as deputy minister or higher. Moreover, because we have a defined population of elites from which we are sampling (see the recruitment protocol discussed above), we can also formally test how representative our participants are, along two dimensions. First, how do our participants compare to the complete population of individuals who served in the Knesset from 1996 to the present? Second, how do our participants compare to our sampling frame, a different group than the complete population because it does not include members who had passed away, were too sick to participate, or for whom we were unable to acquire up to date contact information. Thus, whereas the first quantity explores whether our participants look like the universe of Knesset members in this time period, the second explores survey non-response. We explore both questions in Table 3 below, which presents a set of linear probability models comparing our participants to the universe of Knesset members from 1996-2015 (models 1-2) and to only those Knesset members who had been sent the survey (models 3-4). The results show that unsurprisingly, current members of the Knesset were less likely to participate in the survey than former members, but that interestingly, our participants are not significantly less “elite”, as measured by the proportion of respondents with experience as deputy ministers, or as cabinet members or higher. If anything, our sample is slightly more experienced than the universe of decision-makers, though the number of terms in office did not significantly predict survey response.

Table 2: Knesset Sample Characteristics (N=89)

		Proportion of respondents	
Knesset Member:		Current	25%
		Former	75%
Exp. on Foreign Affairs/ Defense Committee ...			
... as backup or full member			67%
... as full member			54%
Highest level of experience:			
... not a Minister			58%
... Deputy Minister			29%
... Cabinet Member or higher			12%
Male			84%
Served in military			95%
Active combat experience			64%
		Mean	SD
Age		61.4	10.7
Terms in Knesset		3.0	2.1
Military Assertiveness		0.61	0.20
Right Wing Ideology		0.45	0.24
Hawkishness (Arab-Israeli conflict)		0.39	0.25
International Trust		0.40	0.26

Note: individual differences in bottom four rows scaled from 0-1.

3 Content Analysis

Figure 2 in the main paper estimates the proportion of reputational language in daily newspaper coverage of the crisis in Crimea (from January 15-July 15, 2014) and Syria (from July 1-December 31,

Table 3: Sample representativeness tests

	Compared to...			
	All Knesset members (1)	(2)	Sampling frame (3)	(4)
Current member	-0.043 (0.045)	-0.049 (0.057)	-0.210*** (0.054)	-0.184*** (0.065)
Highest level of experience:				
...Deputy minister	0.017 (0.054)	0.044 (0.071)	0.035 (0.072)	0.079 (0.088)
...Cabinet member or higher	-0.044 (0.076)	-0.098 (0.098)	-0.075 (0.093)	-0.096 (0.114)
Male	0.025 (0.053)	0.081 (0.063)	0.072 (0.067)	0.097 (0.076)
Terms in office	0.011 (0.012)	0.021 (0.016)	0.008 (0.015)	0.013 (0.018)
Left-right party membership		-0.070** (0.030)		-0.063 (0.038)
Constant	0.177*** (0.054)	0.312*** (0.087)	0.320*** (0.070)	0.436*** (0.108)
N	415	295	288	225
R ²	0.007	0.043	0.063	0.080

*p < .1; **p < .05; ***p < .01

2012), superimposed by a loess smoother (in blue) On average, reputation, resolve, honor, and credibility are discussed in 9.7% of all articles about Crimea, and 10.5% of all articles about Syria during these six-month windows. For Crimea, the denominator here consists of 49248 news articles available on LexisNexis (from January 15-July 15, 2014) mentioning (Crimea|Ukraine)&&(crisis|war|uprising|revolt|conflict|violence|troops|rebellion|rebels|annexation); for Syria, the denominator consists of 37937 news articles available on LexisNexis (from July 1-December 31, 2012) mentioning Syria&&(crisis | war|uprising|revolt|conflict|violence|violent|troops|rebellion|rebels|revolution|terror). In both cases, the numerator retains articles containing the words reputation|resolve|honor|honour|credible|credibility|credibly, omitting references to phrases found to be unrelated to the quantities of interest (resolved, resolves, to resolve, honorable, honourable, an honor, an honour, honoring, honouring, honorary, honourary, honorific, honourific, incredible) after manually inspecting a random subset of the search results.

References

- Davidov, Eldad. 2009. "Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective." *Political Analysis* 17(1):64–82.
- Davidov, Eldad, Bart Meuleman, Jan Cieciuch, Peter Schmidt and Jaak Billiet. 2014. "Measurement Equivalence in Cross-National Research." *Annual Review of Sociology* 40(1):55–75.
- Duckitt, John and Chris G. Sibley. 2010. "Personality, Ideology, Prejudice, and Politics: A Dual-Process Motivational Model." *Journal of Personality* 78(6):1861–1894.
- Feldman, Stanley. 2013. Political Ideology. In *The Oxford Handbook of Political Psychology*, ed. Leonie Huddy, David O. Sears and Jack S. Levy. New York, NY: Oxford University Press.
- Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1):25–46.
- Huff, Connor and Joshua D. Kertzer. 2018. "How the Public Defines Terrorism." *American Journal of Political Science* 62(1):55–71.
- Jost, John T., Christopher M. Federico and Jaime L. Napier. 2009. "Political Ideology: Its Structure, Functions, and Elective Affinities." *Annual Review of Psychology* 60(1):307–337.