Oxford Research Encyclopedia of Politics

Signaling in Foreign Policy

Erik A. Gartzke, Shannon Carcelli, J Andres Gannon, and Jiakun Jack Zhang

Subject: World Politics Online Publication Date: Aug 2017
DOI: 10.1093/acrefore/9780190228637.013.481

Summary and Keywords

Costly signaling offers a solution to many foreign policy dilemmas. Though most commonly studied in the context of the bargaining theory of war, signaling can also play an important role in nonzero-sum interactions such as those characterized by chicken (e.g., nuclear deterrence) and the prisoner's dilemma (e.g., tariff reductions). A rich game theoretic literature explains how actors can signal credibly in these situations. The most prominent strategies are sinking costs (actions that are costly ex ante) or tying hands (actions that are costly ex post). These strategies are theoretically elegant but have generated considerable controversy when studied empirically. One controversy concerns the existence of hand-tying domestic audience costs under different regime types. A second controversy involves the degree to which sinking costs increase or decrease the risk of war. These controversies speak to the inherent tension between theories of strategic interactions and measuring their outcomes in the foreign policy process, where some events are off the equilibrium path and thus unobserved.

The limited availability of foreign policy data was a major hindrance in earlier empirical efforts. Even as the quality of this data has improved, focus has been on the outcomes of conflict (crisis onset, escalation to war, victory, defeat) rather than the strategy. This is problematic given that all crises are sequential in nature and understanding the action–reaction cycle is vital to illuminating patterns of war, capitulation, and settlement. The frontier of research in the signaling literature is in bridging this gap. The advent of big data and machine learning has enabled more systematic empirical analysis of strategic moves by various foreign policy actors, including signaling. Some researchers, such as Lindsay & Gartzke, are harnessing these new data and methods to explore the means of signaling. Other scholars are beginning to ask questions about the efficacy of public versus private signaling, the role of ambiguity, and dyadic versus multi-actor signaling. This new wave of research seeks to nudge signaling closer to the concerns of foreign policy practitioners.

Keywords: signaling, foreign policy, credibility, communication, bargaining, threats, resolve, tying hands, sinking costs

Introduction

Signaling plays a crucial role in nearly all aspects of foreign policy and has long been the subject of practitioner interest and academic study. Because most foreign policy outcomes depend on the strategic interaction of multiple actors, the behavior of one actor depends on their expectations about the intent and capabilities of others. But in both zero-sum interactions, such as bargaining over territory, or nonzero-sum interactions, such as negotiating a trade agreement, actors face problems of incomplete information and credible commitment because they each have incentives to misrepresent their own capabilities and resolve while communicating with other players to get a better deal (Fearon, 1995). Signaling offers a solution to these common foreign policy dilemmas. There are different types of actions one can take to signal. A thorough review of different characteristics of signaling is technical and is discussed only partially here.

Signaling is most studied in the context of the bargaining theory of war, where both actors are better off with a negotiated settlement rather than paying the costs of fighting. But signaling is also essential in understanding the dynamics and outcomes of nonzero-sum interactions such as chicken (e.g., nuclear deterrence) and prisoner's dilemma (e.g., tariff reductions). A rich game theoretic literature explains how actors can signal credibly in these situations. The most prominent strategies are sinking costs (actions that are costly ex ante) or tying hands (actions that may become costly ex post). These are theoretically elegant but have generated considerable controversy when studied empirically. One such controversy (Fearon, 1997) is over the existence of hand-tying domestic audience costs under different regime types. A second controversy is the degree to which sinking costs increase or decrease the risk of war (Slantchev, 2005). There is also a tension between theories of strategic interaction and measurement of their outcomes in foreign policy processes, where some events are off the equilibrium path and thus are unobserved.

The frontier of research in the signaling literature has shifted beyond the logic of credible signaling and toward theorizing about its practice in foreign policy. Scholars are beginning to ask questions about the efficacy of public versus private signaling, the role of ambiguity, and dyadic versus multi-actor signaling. Some researchers are also taking the means of signaling more seriously. Indeed, while much of the theoretical advances in the signaling literature have been in formal theory with rather limited empirical analysis (Fearon, 1994), the research frontier looks likely to take a more empirical turn. The limited availability of foreign policy data was a major hindrance to earlier empirical efforts. Even as the quality of this data has improved, scholarship has focused on explaining the outcomes of conflict (crisis onset, escalation to war, victory, defeat) rather than on the strategy of conflict. This is problematic given that all crises are sequential in nature and understanding the action–reaction cycle is vital to illuminating patterns of war, capitulation, and settlement (George & Smoke, 1974). The advent of big data and machine learning has enabled more systematic empirical analysis of strategic moves by various

foreign policy actors, including signaling. For example, whether a state uses seapower or another medium to impose costs on an adversary should send different signals of resolve. Gartzke et al. (FORTHCOMING) have argued that the means of signaling matter as much as their intended ends. This new wave of research seeks to nudge the theory of signaling closer to the concerns of foreign policy practitioners.

Defining Signaling

Signaling occurs when one actor knows something of relevance to another actor's decisions (Morrow, 1999). If the decisions of the uninformed actor affect the informed actor, the informed may wish to signal its information to the uniformed. Under this framework, signaling is a rationally calculated decision to alter strategic interactions between actors. A more nuanced definition is that *signaling is the purposive and strategic revealing of information about intent, resolve, and/or capabilities by an actor A to alter the decisions of another actor B to improve the chances that an outcome desired by A is reached when the desired outcomes of A and B are dissimilar*. This definition incorporates Morrow's idea that signaling is about the communication of information and it occurs when that communication adds something new to the repository of the signal's recipient.

Signaling must be purposive, meaning that the recipient interprets a message from the signal. If the recipient does not receive a message, then no signaling occurred (Cho & Kreps, 1987). Miscommunication occurs when the sender and the recipient differ about the meaning of the signal, but it was still purposive in the sense that the intent to transmit a signal and receive the signal were there for the respective actors. This definition also clarifies what Morrow means when he says that signaling happens when there is information that is of relevance to actor B.

Signaling must also be strategic, meaning that it must change the payoffs of the actors involved. There is often much information that is relevant to the decisions of an actor that others may not wish to communicate. During a conflict, actor A may know that actor B's troops are vulnerable to an aerial strike and will be defeated quickly if they do not move, but actor A may be unwilling to communicate this information if A and B are on opposite sides of the conflict. Communicating credible information about B's vulnerability would affect the informed actor but in a way that produces an outcome that is less desired by actor A. Morrow would likely agree with this definition since he said that "the latter *may wish* to signal its information" (emphasis added). However, it is still conceptually useful to clarify that an actor chooses to signal when it believes that the knowledge of that information by actor B means that actor B is more likely to choose an action that is preferable to actor A than whatever is expected to occur otherwise (Morrow, 1999).

International relations is replete with examples of signaling that span a wide variety of subject areas. Different foreign policy issues dictate the nature of the strategic interactions that require signaling. Foreign policy issues can be broadly construed as being either conflictual (zero-sum), where gains for one side mean losses for the other, or cooperative (nonzero sum), where the sum of the gains and losses for all sides is positive even if they compete over the distribution. The second dimension across which foreign policy issues can vary is the number of actors involved; they can be bilateral foreign policy decisions where the action has a specific initiator and recipient (e.g., diplomatic threats, economic sanctions, military operations, foreign aid) or multilateral, where the

strategic interaction has foreign policy consequences for a number of actors (tariff reduction, human rights, international law, climate change, etc.) Figure 1 shows the typology of the types of foreign policy issues and the primary role that signaling can achieve in each type.

Table 1. Typology of foreign policy issues and the role of signaling

	Multilateral	**Bilateral**
Conflict/zero sum	II) Signaling alliance commitments	I) Signaling to demonstrate resolve
	E.g., extended deterrence (Fuhrmann & Sechser, 2014)	E.g., bargaining model of war (Fearon, 1997)
Cooperation/ nonzero sum	III) Signaling to solve collective actions problems	IV) Signaling willingness to cooperate
	E.g., coordination games in macroeconomic policy (Hall & Franzese, 1998)	E.g., prisoner's dilemma in trade liberalization (Keohane, 1986)

In nonzero sum foreign policies (cells III and IV), purposive communication can help overcome collective action problems and achieve cooperation. This can be in a bilateral setting where a concession is made (the purposive signal) to encourage the other side to make a concession in return (the strategic interaction). For example, in trade liberalization, states are wary of reducing tariffs unilaterally because domestic industries would suffer from foreign competition unless tariff reductions are reciprocated. As a result, even though free trade could benefit both economies, states remain protectionist. By signaling their willingness to cooperate as long as trade liberalization is reciprocated, states could adopt a tit-for-tat strategy of tariff reductions in order to overcome the one-shot prisoner's dilemma equilibrium (Keohane, 1986). Much of the recent literature in international political economy has built upon Milner (1997), who showed that transparent domestic politics can help democratic leaders signal resolve in international negotiations over tariff reductions, allowing liberal regimes better to cooperate.

Signaling could also enable cooperation in a multilateral setting, where the increased number of actors contribute to collective action problems. Signaling can help solve time-inconsistency problems by offering a way to credibly commit to cooperation. One area in which this plays out is in macroeconomic policy coordination. Governments want to keep inflation low, but they are also tempted to use fiscal policy to boost employment. Here the independence of central banking can play a critical role in credibly signaling the desire to cooperate with other actors (domestic and foreign), ensuring the success of monetary policy (Hall & Franzese, 1998). States can also use signaling to establish a focal point to

overcome coordination problems by setting standards, writing laws, or establishing institutions. A significant portion of diplomacy involves this type of signaling—purposive communication that helps states realize that their interests are aligned on issues such as technical standards, climate change, or human rights. Although signaling can help achieve international cooperation by changing the payoffs by unlocking potentially unrealized gains, this type of signaling is less studied by political scientists, even though ironically they are less strategically challenging. The challenges of cooperation (the tragedy of the commons, free-riding) cannot be solved by cooperative signaling alone but instead require other actions, such as sanctioning, which themselves serve as noncooperative signals of ability or resolve.

Traditionally, signaling in international relations is studied in the context of conflictual foreign policy. Signals are often discussed in terms of carrots and sticks, where one state uses incentives or sanctions to encourage a change in the behavior of another state (Bemelmans-Videc, Rist, & Vedung, 2011). The behaviors of interest can be foreign policy decisions that aim to maintain the status quo (often referred to as “deterrence”) or to change it (“compellence”), but the most relevant distinction is whether actors’ goals are fundamentally in tension with one another or whether they are largely complementary. Signaling plays a particularly important role in these zero-sum settings (cells I and II) because what is revealed to the adversary not only changes his payoffs but also can change one’s own. In other words, this type of costly signaling differs from “cheap talk” in that it must change the payoffs of the actors involved. Costly signaling appears of greatest interest to both practitioners and scholars, as much more is at stake when signaling fails.

In foreign policies that are zero sum in nature, such as bargaining over territory, signaling can help avoid bargaining failure and the subsequent move toward war. This matter is well studied in the context of a two-player game described by Fearon, who also investigated how domestic politics can affect the credibility of signaling at the interstate level. Other research has extended this framework and the importance of signaling to multiplayer bargaining games. This is most notable in work on extended deterrence where a “defender” state can signal resolve to protect a “protégé” through forming an alliance or deploying forces to overseas bases. These foreign policy decisions are a means of signaling the resolve to protect the protégé from attack by potential aggressors.

date: 29 August 2017

First Wave: When and Why Do States Signal?

The first wave of research on signaling was concerned with understanding process and with defining the scope and scale conditions for the phenomenon. Given that the present focus is on signaling that happens in targeted-conflictual situations, emphasis on signaling in the context of war is in order. Schelling (1960) described signaling as part of international bargaining and argued that limited war was a form of signaling because actors would anticipate the potential for (possibly unintended) escalation during smaller disputes. As a result, war itself could be considered part of the bargaining process since threats, proposals, offers, and concessions during a military contest are designed to change the opponent's mind regarding their level of commitment (Schelling, 1966). Here, signaling informs because states must pay for costly action in order to communicate resolve to others. Schelling demonstrated that when actors differ in their value for issues in dispute, they can inform by being willing to bear the costly burden of a contest. Small contests are less informative, but also cheaper and easier to initiate. Often, a cheap contest can substitute for a more expensive one, particularly in the nuclear era.

Taking this logic one step further, contests could also be won by demonstrating a willingness to experience a given cost, even if the cost was not incurred. Competition then centers round an actor's relative willingness to incur cost and on the perceptions of other actors about this willingness. One could win a dispute without force by faking resolve (bluffing) or alternately by taking risks of escalation that an adversary finds prohibitive. The chicken metaphor for this type of competition emphasizes both elements of a zero-sum struggle and the coordination that is needed to avoid a mutually harmful outcome. Counter to considerable intuition in diplomatic affairs, Schelling (1966) emphasizes the value of reducing one's options. "If I cannot swerve, and my opponent knows this, then she or he must act with discretion." In the extreme, brinkmanship crises are based on "a strategy that leaves something to chance," where credibly threatening mutual annihilation is not credible. Signaling is the only technique that brings the enormous power of nuclear weapons to bear on political competition, without requiring their use.

Finally, reputation is a further valence on force, based on perceptions of capability and resolve, which in turn are driven by previous success in signaling. During the Cold War, the perception of allies and adversaries that the United States was willing to use nuclear weapons to defend Europe from Soviet incursion allowed it to maintain considerable influence with a much smaller conventional force, freeing up resources for other, more peaceful, purposes. In this strategy, it was critical that Soviet leaders perceived that the United States was willing to escalate first to nuclear use and that it was willing to do so in response to any significant incursion into NATO territory. Reputation was enhanced in a number of ways, including U.S. and allied behavior during a series of brinkmanship crises in the 1950s and 1960s. Reputation is thus a "stock" of perceptions, while signaling itself is a "flow," which can contradict reputations. The relationship between short-term

influences on perception (signaling) and longer-term perceptions (reputation) has not been extensively analyzed in international relations, but it relates to the notion of dynamic rational updating (Bayes's rule) in statistics and economic theory.

Fearon (1995) further examines the relationship between the first two of these processes outlined by Schelling. The elemental problem posed by conflict is that it appears inefficient. Why would rational states or other actors incur (impose) costly war instead of negotiating settlements that are ecologically efficient and that can be made at least as beneficial individually? Fearon builds on Blainey (1973), who notes that war must reflect different expectations about the outcome of a contest. If I think I am going to win, and you think you are going to win, then no bargain may suffice to make us both satisfied. Obviously, one of us is wrong; war will punish at least one side's preconceptions. Blainey concludes that war is the product of misperception and therefore is fundamentally an irrational process. War ends when adversaries agree on their relative power.

Fearon challenges the notion of the need for irrationality in motivating war. Actors may fail to perceive the same reality simply because one or more prospective combatants are uncertain about relative power or resolve. Fighting then informs in much the same way as both Blainey and Schelling imagine, but also reconciling different conceptions or perception. Signaling is most relevant to the idea of war as alternative to a negotiated settlement ("bargaining failure") because actors have private information and incentives to misrepresent. Fearon notes that conflict does not require both sides to expect to win a dispute, only that expectations of relative performance differ and are incompatible. For fighting to occur, the first opponent has to be more optimistic about its success in battle than the other opponent is, even if both agree that the first opponent will lose eventually. The first may believe that, by imposing unacceptable costs, the second opponent will tire of the contest. Conversely, the second opponent may suspect that the first's optimism will be muted by incremental military defeat.

Each side in a prospective dispute knows best its own capabilities. This "private information" about military factors results in different estimates of the probability of victory. Signaling can help to avoid a costly contest by revealing private information. Again, bluffing is attractive and therefore a problem. If one side can pretend to be more resolved or capable, and be believed, then it can have its way in diplomatic negotiations and also avoid war. Thus, competitors have an incentive to pretend to be capable or resolved, even when they are not. For this reason, most "cheap talk" claims of wartime performance are not believed. Actors must prove military prowess or intent, either by fighting or preparing a proxy for war that is also credible. Countries or nonstate actors can reveal capabilities by exhibiting them, parading them through the streets of the capital city. They can conduct war games or deploy forces to areas of interest. Even the manufacture of military power can prove influential, as indeed is post behavior. All of these actions are signals meant to shape perceptions and influence current or future bargains without necessitating force.

There is a double-edged sword to signaling; while revealing one's "true" capabilities may help avoid an inefficient contest and the costs of war by encouraging negotiation ex ante, it could also compromise efforts to obtain a favorable settlement. Depending on the nature of one's military advantage, revealing capabilities forewarns an adversary and may allow an opponent to better prepare her defense. The very act of signaling one's capabilities may render those capabilities less effective. If, for example, one were to share war plans with an adversary in an attempt to convince the adversary of the quality or veracity of an attack, this might blunt the effectiveness of the attack. A surprise attack is presumably much less likely to prove decisive if it is no longer a surprise. In sum, because signaling in the crisis context is motivated by a desire to achieve a more desirable negotiated bargain, actors have an incentive to misrepresent their true capabilities in ways that make war more likely. The familiar paradox is that a desire to avoid war is itself an inducement to warfare (Wagner, 2000). Similarly, actors that seek to communicate via costly signaling in a crisis, to differentiate themselves from weak types, may heighten the risk of escalation (Slantchev, 2011; Fuhrmann & Sechser, 2014). Given ambiguity about the incentives to signal or conceal one's military actions, what has been called the "spiral model" (Jervis, 1976) suggests that costly signals like military mobilization can be interpreted variously by observers as evidence of the need for compromise or as preparation for war (Fearon, 1995; Sartori, 2002).

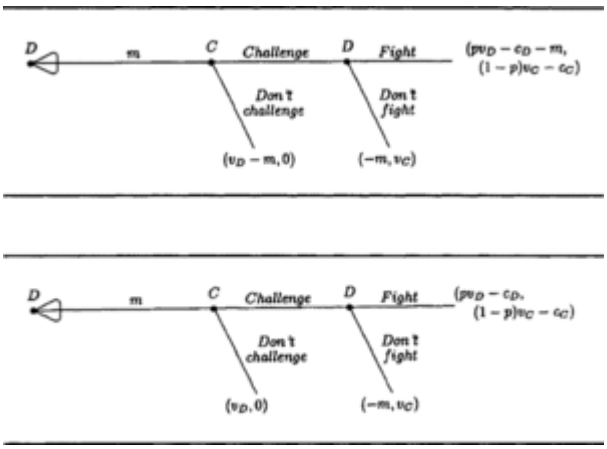
Fearon (1997) suggests two archetypal ways that states can signal, labeled colorfully as "tying hands" and "sinking costs." Sunk costs are incurred when a chosen action is costly up front and when the action has minimal subsequent effects on the choice making of the actor taking the action. Paying the salaries of soldiers deployed as part of a defending force is a sunk cost, both because it is an expense regardless of whether there is an attack and because the soldiers must be paid again next year if one wishes to continue to retain their services. The "sunk" portion of sunk costs is important to isolate for analytical reasons because nonsunk costs can shape the incentives for using force. It is often difficult to separate out these different effects of up-front spending. Mobilizing troops, for example, changes the defender's decision about whether to fight based on the military balance, value for the prize in question, and cost of fighting.

Sunk costs are informative to the degree that they differentiate resolved or capable actors from those that are less willing or able to fight. It is assumed that a defender invests less in signaling its resolve when its costs for fighting are high or its value for the prize is low. Again, to the degree that costs are sunk, they don't affect the defender's actual decision to acquiesce or resist.

A second, contrasting method of signaling involves tying hands. Actors can influence and inform by imposing on themselves a cost that they *only* incur in the event that they fail to act in a manner consistent with their ex ante claims. A state, for example, can form an alliance. The alliance itself imposes few costs, since it is just a "scrap of paper." However, an explicit promise to protect an ally is informative to the degree that it ties that defender's reputation to its future action in a crisis. Failing to honor an alliance commitment can be costly in that other nations will think twice before believing claims by

the defender in the future. States can thus tie their future reputations to actions ex post, achieving more effective deterrence or influence generally ex ante. As a result, the defender and its protégé are in a better bargaining position relative to possible challengers, since the incentives that actor faces in the future have now changed.

Figures 1 and 2 are taken from Fearon (1997). They depict full information games for sunk-cost signals and tying-hands signals, respectively. The figures demonstrate the difference in when costs are incurred on the defender depending on the type of signal that is sent. If the defender sends a sunk-cost signal, like military mobilization, that cost (denoted here as -*m*) is incurred regardless of how the challenger responds. In comparison, the cost of tying-hands signals in Figure 2 (also denoted as -*m*) only exists if there is an actual challenge. If the challenger opts not to challenge, that signal imposes no cost on the defender that issued that signal. Instead, the audience cost present in tying-hands signals "are paid by the defender only if the defender backs down or backs away from a challenge" (Fearon, 1997, p. 73).



Click to view larger

Figure 1. Models of sunk-cost signaling and tying-hands signaling. The models differ as to when the costs of signaling are incurred [Credit line: Fearon (1997). Signaling foreign policy interests tying hands versus sinking costs. *Journal of Conflict Resolution*, *41*(1), 68–90.]

Fearon (1997) concludes his study of alternative signaling mechanisms by emphasizing that so called cheap-talk signaling (tying hands) is preferred to sunk costs precisely because tying hands is cheaper. However, as always seems to be the case with strategic interaction, the "better" option may not be the preferred action. The fact that tying hands is likely cheaper means that more actors will choose it, making its signal less exclusive and therefore less informative than sinking costs. An actor intent on demonstrating her willingness to fight may want to choose to sink costs simply because other, less resolved actors will prefer tying hands instead. In technical terms, costs sunk in a signal may do more to "separate" types, providing better information.

Similarly, the differences in the two mechanisms are evocative of differences in their utility. By its very nature, sunk-cost signaling does not impact future choice making by the signaling actor. Sunk-costs signals thus operate on "credibility," the component of signaling that has to do with reducing uncertainty about an opponent's attributes. This is presumably particularly desirable for actors that (1) are highly resolved or capable, but (2) are not perceived to be by other actors. In contrast, cheap talk signaling is intended to change the incentives and thus subsequent behavior by the signaling actor. Tying hands

therefore operates more through “commitment,” the mechanism meant to inform observers of a *change* in capability or resolve, whatever the initial level of these factors. Tying hands presumably matters more for actors that are not resolved but would like to be, or that are moderately resolved but want to be more so. Joining a dieting club may do little for someone who is actually going to lose weight no matter what happens, but it could be an effective way for people to increase their resolve, if they are concerned that old habits may prevail. Conversely, the purchase of expensive exercise equipment will not cause someone to exercise but telegraphs the intention to become fit, if in fact one is already resolved.

Finally, even if, contra Fearon, war is thought to be *part* of the bargaining process rather than an *alternative* to it, signaling still plays a similar role in helping states get closer to an actor’s desired outcome. Wagner (2000) argues that war is itself integral to the bargaining process since most wars end with an agreement rather than being a “fight to the death.” In this case, war can be seen as a signal because it refines actors’ prewar expectations about the likelihood of victory, which determines the potential range of negotiated settlement options. Wars serve the function of revealing information about resolve and/or capabilities, which explains why exchanges of offers between warring states are not constant; war allows actors to update their expectations about the relative utility of continuing to fight versus accepting a negotiated outcome. Wagner’s refinement of Fearon’s basic model of war indicates that because a costly contest can be integral to the bargaining process, fighting can then be seen as a costly signal designed to achieve a desired negotiated outcome. Either explanation is consistent with the idea that an actor strategically signals its intent, resolve, and/or capabilities to improve the chances that an adversary’s actions are more consistent with the actor’s desired outcome. Thus, the definition of signaling given earlier represents cases of where signaling occurs prior to conflict to avoid it as well as cases where signaling occurs during a contest in an effort to update beliefs about resolve and capability.

Psychological explanations of deterrence also seek to explain signaling. Jervis (1989) builds on Schelling to argue that the perceptions of decision makers about the world diverge from reality in predictable patterns. This divergence explains conflicting predictions from the deterrence and spiral models of conflict. The deterrence model argues that when aggressors believe status quo powers are weak in their capability or resolve, war becomes more likely. States then have to go to extremes to avoid actions (moderation or conciliation) that are seen as weakness. As a result, aggressors will try to challenge a status quo power if the aggressor thinks the status quo power is weak (Jervis, 1989). In contrast, the spiral model holds that states act on worst case assumptions and perceive security-oriented actions by others to be a threat to their own security. The result is a spiral in which each attempt to improve one’s security is met with a response that threatens the actor’s security. Jervis’s innovation is this psychological explanation for conflict that looked at actors’ perceptions of themselves, other actors, and the environment.

It is important to recognize, however, that conflict spirals need not follow from the psychological framework that Jervis provides. First, unnecessary escalation can stem from rational uncertainty, as outlined by Fearon (1995) and as addressed in subsequent research. Mobilizations can deter or inflame, depending on one's interpretation. Interpretations might be entirely defensible in ecological terms. Sometimes an enemy is trying to mount a surprise attack. At other times, they may be posturing for a better situation. Second, a remedy for the psychological model exists in the model. A decision maker learning about the nature of spirals from Jervis or other scholars can take this phenomenon into account. Aware of an adversary's misperception, a target can steer clear of controversy and act with greater discretion (Leeds & Davis, 1997). The net result of this insight, however, is to reduce the empirical connection between perceptions and war. As with many social phenomena, knowledge of the phenomena weakens the reported effect. From this insight can also evolve a better understanding of strategic communication more generally.

Second Wave: When Is Signaling Effective?

Much of the first wave of signaling research either predicted or assumed that signaling is an effective tool of economic or military statecraft. The second wave of signaling research clarified at least three main issues in this regard. First, the theoretical value of certain signals may not be empirically justified. In the second wave, scholars began specifically questioning the theoretical microfoundations of foreign policy signaling, especially the domestic-policy consequences of backing down. Second, states can use more than actions to signal. In particular, second-wave scholarship has focused on institutions. Certain institutional setups (such as an opposition party) can make signals more credible. Because states are not unitary actors, various domestic interests can independently signal the resolve of the state as a whole. Third, the credibility of a signal is not all that matters. Different types of signals may have different types of consequences, and heterogeneity matters in ways we are still trying to understand. Second-wave researchers looked more specifically at how and why a signal may or may not be effective. In doing so, they have begun to seriously consider heterogeneity in the means and ends of foreign policy signals.

In a game theoretic formulation, signaling in an international crisis should be effective insofar as it allows the signaler to separate him- or herself from less resolved types. This requires that signals be relatively costly. In Fearon's original models, neither tying hands nor sinking costs allow for the possibility of bluffing. States show their resolve, and their opponents correctly interpret the signals. This advances an optimistic assessment of the value of signaling in foreign policy.

However, as Slantchev (2005) suggests, a signal of a given cost may fail to differentiate between different types of signaling actors. In fact, weak or unresolved states have incentives to bluff by undergoing costly military mobilizations. In doing so, states can simultaneously sink costs and tie their own hands. Military mobilization is both a signal and a straightforward attempt to win an upcoming war. Unlike either tying hands or sinking costs, military threats *both* are inherently costly (due to audience costs) *and* change the (local) distribution of power. They allow states to restructure their own preferences in a way that increases the probability that they will win if a war breaks out. Because of this, military mobilizations create a possibility that a costly war will break out despite states' initial intentions to avoid war. This perspective offers a more pessimistic or possibly mixed assessment of the effects of foreign policy signaling on crisis outcomes.

Audience Costs

Tying one's own hands by making military threats is valuable only if leaders face penalties for breaking their promises. These penalties may be felt through an international audience (through reputation) or a domestic one (through loss of power).[1] It is easy to imagine a world where leaders face penalties for backing down in a conflict, especially among democratically elected leaders who depend on the trust of their constituents. However, the precise mechanisms leading from (1) a leader backing down after promising to fight to (2) the leader being punished in some way are not well developed. Further, the empirical association between making empty threats and facing international or domestic retribution is hazy. The concept of audience costs lives on, but scholars continue to debate and study the precise theoretical mechanisms and empirical effects of audience costs (Gartzke & Lupu, 2012; Schultz, 2001A; Snyder & Borghard, 2011). The literature on audience costs is extensive and is growing rapidly, with new methods being applied and old debates continued. Our treatment of audience costs here is cursory; the subject is addressed extensively elsewhere in this volume (Hamilton, 2017; Ramsay, 2017; Weeks & Crunkilton, 2017).

Fearon (1994) provides the initial intellectual impetus for audience cost research and defines the term. However, there are peculiarities in the model he offers, reflecting debatable assumptions about the motives of domestic political audiences. Smith (1998) provides a much more general mechanism of domestic audience costs by endogenizing the decision to make a threat and back down. Rather than treating voters as if they inherently prefer honesty, Smith emphasizes efficacy. He argues that incompetent leaders, who know they will be punished electorally regardless of their actions in any particular crisis, will be the most likely to make a threat and back down. This in turn signals to constituents that the leader is the incompetent type. Therefore, Smith expects that audience costs will only distinguish the least competent type of leader, while leaders of intermediate or better competence may be baited into a war they do not desire.

As with any novel and thought-provoking insight, it is the accumulation of context and nuance that serves to enrich and diversify effects. Politics is an adversarial game. Initial audience cost mechanisms had no explicit role for the opposition. Domestic political opponents could serve as a mechanism to mobilize the public and either augment or downgrade the content of international signals. According to Schultz (1998), domestic strategic opposition parties reduce the probability of war by enhancing the quality of information about national preferences. Threats become more credible when the opposition publicly supports the government. This is partly a result of the fact that knowledge that the domestic opposition accepts government policy ensures social cohesion and lower domestic political costs. But it also suggests that the government correctly identified the "national interest," implying competence and effectiveness to the foreign power. In cases where the opposition party stands up to the government, threats can become more credible because the domestic risk of bluffing forces the government to choose threats selectively.

Other domestic institutions can serve to increase the credibility of signaling. The "fourth estate" can operate much as a domestic opposition group, making it costly to bluff and hard to resist the will of the population (van Belle, 1997). Slantchev (2006) proposes a similar mechanism in which certain institutional settings allow the media to act as an intermediary between leaders' actions and audience costs. The courts may also operate to increase the credibility of signals in foreign policy, though at least in the United States, the courts have been reluctant to weigh in on issues of national security and generally leave foreign policy to the executive branch.

At the same time, it is far from clear that third parties necessarily operate in ways that invariably clarify rather than obscure. The very nature of politics encourages occasional obscurity. Just as Smith (1998) and others find fault with a nonstrategic electorate, others have reason to quibble about the characterization of the motives imparted to the domestic opposition. As Ramsay (2004) demonstrates, the circumstances of domestic politics may impose themselves on foreign affairs. A leader with a "safe seat" may receive support from an opposition group that does not wish to lose additional ground with the public, while one that could lose the next election may be challenged, if for no other reason than this improves the electoral prospects of the opposition. "Gambling for resurrection" can occur at different political levels (Downs & Rocke, 1995). A similar perspective is offered by Trager (2002), who sees coalitional politics as a critical variable.

Nor is it clear that audience costs always operate for identical ends. As implied by this discussion, leaders may generate audience costs in several ways. Signaling in the face of domestic opposition can demonstrate resolve, while support from putative domestic opponents indicates greater credibility for a leader's claims that he or she is backed by the nation. Obviously, the former precludes the latter and vice versa. Precisely because it is easier to threaten abroad when one has domestic allies, it is less likely that the leader will pull his or her punches and less costly if he or she has to back down. Domestic support can thus strengthen or weaken the import of a leader's demands, depending on what is being asserted and what conditions entail. As with any other form of signaling,

the effect can be directed more toward commitment or credibility, and the impact of the signal may be used as leverage in bargaining as much as or more than a tool to reduce conflict.

Empirical research on signaling through audience costs has been less than conclusive. Early efforts focused on implications that, while plausible, were not exclusive to what is after all an intangible process (Partell & Palmer, 1999; Schultz, 2001A). Critics doubt that audience costs even exist, based on their reading of historical crises (Snyder & Borghard, 2011; Trachtenberg, 2012). Advocates are quick to point out that audience costs are not transparent in research terms, based on some of the factors outlined earlier (Schultz, 2001B). However, the claim that a theory is not testable is hardly ever satisfying, and efforts have been made to better assess audience cost signaling in ways that make it possible to observe and infer causation. At the same time, the notion of audience cost has been applied ever more broadly, including now autocracies, which initially were thought to be the contrasting (nonaudience cost) case (Weeks, 2008; Weiss, 2013).

A potential solution to the problem of unobservable mechanisms has been to create hypothetical scenarios and test them experimentally. Tomz (2007) and, later, Levendusky and Horowitz (2012) created survey experiments in which they manipulated the actions of hypothetical leaders. Research has found that voters are likely to punish leaders who back down after making a threat, but that leaders' explanations for backing down may ameliorate those effects. Tingley and Walter (2011) conducted a similar experiment on international audience costs, asking whether "cheap talk" can deter in an experimental setting. Although the research design brings us closer to seeing the mechanisms behind audience costs, they suffer from external validity challenges. Because study participants rarely see a leader backing down in real life, they may pay more attention than they would were it a more common occurrence. It therefore may not be the act of backing down that participants are responding to, but the novelty of the hypothetical event itself.

One area where audience costs have proven promising empirically is in the role of international institutions. Theorists have argued that international institutions can serve a special validating function in the signals of state leaders (Thompson, 2006; Chapman, 2009). Experimental work on international audience costs in particular has demonstrated that respondents value the opinions of intergovernmental institutions such as the United Nations and NATO in assessing the validity of their leader's threats to using force (Tago & Ikeda, 2015; Suong et al., 2017). Questions remain about the causal logic for these relationships (Chu, 2017), as well as the likely strategic impact of international organizations' audience costs on the decision of leaders to signal, but this is a promising area of research.

Reputation

Signaling is closely tied to the notion of reputation in state affairs. In fact, reputation is a key variable for most processes in international relations. Along with allowing states to signal their intentions in a crisis, reputation can be powerful in cooperative bargaining as well. States' past behavior may come back to haunt them, which can affect their credibility in both war and peace. Guisinger and Smith (2002), building upon previous work by Sartori (2002), suggest that leaders benefit internationally from a reputation for honesty. However, empirical research has shown mixed results on the importance of reputation. Huth and Russett (1984, 1988) and Huth (1998) find little empirical justification for a link between reputation and success in extended deterrence situations, although they identify some importance for previous bilateral encounters.[2]

Both the methods and findings of work on reputation have been criticized. According to Mercer (1996, 1997), the value of reputation has been misunderstood largely owing to scholars' insistence on rationality. Mercer highlights the emotional value of reputation, arguing that context and attribution can be powerful factors in states' valuations of reputation. Press (2005) examines the role of reputation—measured by states' behavior in previous crises—in their future credibility. He argues that past behavior is a weak signal compared to other signals of power and interests. Just because states backed down in the past does not mean they are likely to do so in the future. New contexts and issues can make past behavior seem obsolete. Responses to this work have more carefully considered the role of context. Crescenzi (2007) agrees that states must consider contextual clues when evaluating their opponent's reputation in a crisis and finds that states are indeed likely to consider the context of previous crises before attributing resolve. Past behavior against similar adversaries will be more influential than very different types of events.

The causal effect of reputation is especially difficult to measure empirically, partly because it is so important to leaders. Past actions and signals condition the present and the future; common conjecture about who is likely to prevail in a crisis is baked into leaders' expectations, causing them to accept (often silently) conditions that they don't prefer (Cetinyan, 2002). As Huth and Russett (1990) and, more recently, Gartzke and Lupu (2012) point out, leaders who back down in a crisis may be unwilling to acknowledge that they backed down. This makes it difficult for researchers to measure whether or not the state in question actually made a threat. Far from indicating that reputation does not matter, the presence of this selection bias shows that it may matter quite a lot. Similar dynamics may play out for domestic audiences. As Schultz (2001B) suggests, leaders are aware of these costs when they make a threat in the first place. Therefore, many of the potential mechanisms of audience costs are unobservable.

Empirical work has responded to this challenge by considering reputation outside of the crisis context. Gibler (2008), arguing that alliance formation does not suffer from the selection problems that plague the crisis bargaining literature, considers reputation in international cooperation. He finds that states are hesitant to form alliances with partners that have violated previous commitments. Crescenzi et al. (2012) also focus on the alliance literature, with similar findings. Because an understanding of reputation is valuable both in war and in peace, scholarship drawing from different areas of international relations literature have provided theoretical and empirical insight into the actual importance of reputation as a costly signal.

Experimental work applied to foreign policy signaling must demonstrate that the individual unit of analysis matters. So far, there are no broad theories that make predictions at the individual level that these experiments are able to test. Kertzer and Brutgar (2015) begin to address this matter by separating inconsistency from belligerence in an experimental setting.

The Third Wave: Heterogeneity in Means and Ends

States have a powerful toolbox at their disposal to signal their intentions. Since Jervis's (1978) discussion of offensive and defensive power, scholars have been considering how systematic differences in intentions and capabilities affect state behavior. Foreign policy signals can vary in several dimensions, including their costs, their value for winning an immediate contest versus warning an adversary away from escalation, and their credibility over time. The deterrence literature generally assumes away many of these differences. For example, scholars tend to reasonably assume that nuclear threats are not credible for preventing a conventional war (e.g., Powell, 1990) and that conventional threats cannot deter a nuclear attack (e.g., Waltz, 1981). More recent research is beginning to consider the mechanisms behind these broad, intuitive differences. However, we still know very little about the substitutability of signals in different domains: conventional, nuclear, diplomatic, economic, or social (Handberg, 2016).

Previous work began with the development of theories on the differences between weapons systems within the conventional and nuclear domains. Mearsheimer (1983) suggested the certain types of conventional weapons, such as precision-guided munitions, can change the shape of warfare and therefore the calculations that states make before war breaks out. The differential signaling values of these weapons, as well as their importance relative to other diplomatic and weapons signals, remain to be understood. Recently, scholars have begun making similar comparisons across weapons systems in the nuclear domain. Gartzke et al. (2014) find that capabilities differ depending on a state's nuclear force structure. Nuclear states tend to diversify their portfolios over time in a

predictable manner. Once again, little remains known about the signaling value of various elements of states' nuclear arsenals (Gartzke et al., 2017).

Scholars also continue to debate the value of diplomatic signals, owing to their perceived low cost. As mentioned previously, costly military signals may be more powerful than "cheap talk" because they allow resolved states to differentiate themselves from other, less resolved, actors. However, unlike costly threats, diplomacy is unlikely to lead directly to unintentional war (barring the audience cost mechanisms mentioned earlier). Truly resolved leaders may tie their hands, sink costs, or do both simultaneously through military threats or mobilization (Slantchev, 2005, 2010). The means by which a state signals its intentions is a reflection of both capabilities and an influence on the nature of the actions to come. Even when states are initially inclined to favor peace, an attempt to signal through a costly military mobilization may lead to war because such signaling serves to reshape the incentives of both the sender and the receiver.

Recent work has further established the idea that different types of signals can have differential effects on crisis outcomes. For example, Lai (2004) differentiates between public versus private mobilization, showing that *how* states mobilize their forces is as important as *whether* they choose to do so. States that mobilize privately have perhaps intentionally failed to signal their intentions as powerfully and thus are more likely to face the prospect of war. States that have a higher expected utility for war will employ mobilization strategies that are more likely to lead to war because they are designed to maximize the tactical effectiveness of forces. Lai concludes that war is more likely when a state undertakes a mobilization strategy designed to maximize its war-fighting ability. Gartzke et al. (FORTHCOMING) shows that different naval technologies, based on their visibility, have differential values for winning (diplomatic influence, power projection) versus warning (dispute propensity) in a crisis. This work suggests that further differentiating types of political tools (military, economic, diplomatic) may lead to further insights into the role of signaling in war and peace.

Other recent research has vastly improved our understanding of signaling on a number of dimensions. While initial work on signaling was inspired by economic models of signaling games that involved a sender, receiver, message, and reaction, the international system often includes many more actors (Cho & Kreps, 1987). The intended recipient is not the only actor affected by the intended signal that is sent. For example, nondyadic models of signaling have investigated the role of alliances (Fuhrmann & Sechser, 2014). In the context of foreign nuclear deployments, verbal alliance commitments are credible when they are public since the reputation of the state making the promise is put on the line. This serves a hand-tying function similar to that conceptualized in the past, but this signal is more effective than previously thought. Nuclear forces stationed on allied territory only reduce the chance that an ally is a target of violence when it is accompanied by a formal defense commitment from the nation stationing its nuclear forces there. This demonstrates that alliances, despite seeming little more than "pieces of paper," do serve as credible signals of commitment because of how they engage one's reputation. Earlier research may have thus underestimated the effectiveness of cheap talk in extradyadic

contexts. The number of players can also alter the effects of signaling by creating more avenues for the transmission of information, thus increasing the set of equilibrium outcomes (Trager, 2015). More actors means additional costless means of communication that are credible because they alter the protégé's conduct. Protégés make a high demand when they receive support from the third party in the form of signaling. Because the third party cannot credibly convince other players that they will only defend the protégé under limited circumstances, the risk of war increases because allies become more assertive after their allies have communicated resolve. Costly or valuable extradyadic ties, such as networks of trade partners, may also play a role in transmitting credible information about resolve (Gartzke & Weseterwinter, 2016).

This work on nondyadic signaling is an example of a broader shift from analyzing the sender and receiver toward analyzing the multiple options available for information transmission. States almost always have a vast array of options for communication during interstate diplomacy (Trager, 2010). Diplomacy is often considered uninformative ("cheap talk"), since it imparts no or very little cost to the sender. However, the medium itself can be the message. According to Trager, the decision to use diplomacy is relevant information in itself, as the act of talking can change perceptions and payoffs. Trager identifies cases in which a resolved state may want to pool with unresolved states to catch their adversaries unprepared. By allowing actors a suite of options, Trager mediates between models of diplomacy (e.g., Sartori, 2002, 2005) and classic deterrence models that are skeptical of verbal communication. As with both conventional and nuclear forces, diplomacy may have a role in signaling states' intentions. Formal models of signaling previously assumed an actor would respond to a signal by either standing firm or conceding, but this glosses over numerous ways that an actor could try to stand firm and different degrees of concessions that alter the strategic calculations of the signaler in a variety of ways.

When signaling was thought of as involving a sender, receiver, and message, the message was understood as mattering only if the receiver knew that it was sent. As a result, signaling was only examined in the context of overt and explicit statements or actions about which the receiver was aware. An outgrowth of the nondyadic research, however, has been an examination of the role of secrecy in signaling. Covert communications were initially understudied because researchers assumed that private signals were not credible and in any event could not be observed. However, new studies of covert Cold War interventions reveal that secretive actions or communications are often at least partially observed by the receiver (Carson & Yarhi-Milo, 2017). When a receiver knows that a signal is hidden from other audiences (domestic audiences or allies, for example), the covert message can inform, and even demonstrate resolve, because the signal is observably constrained by the impact of the signal on nontarget audiences. This work has questioned the conventional understanding of costly signaling as previously conceptualized by Schelling and others. Although private diplomacy does not have a hand-tying mechanism in relation to the sender's target audience, the fact that there is also no signal communicated to the defender's target audience can allow a defender to save face when

they capitulate since no explicit, public threat forces a leader into intransigence (Kurizaki, 2007). Importantly, this shows that leaders cannot simply ignore private threats as less costly, and therefore less informative, signals.

Much has been made of what has been termed the “individual turn” in international relations that examines the behavior and characteristics of leaders and other critical actors to explain variation in decision making and goals (Chiozza & Goemans, 2011; Horowitz & Stam, 2014; Saunders, 2009). In the context of signaling, an important question is whether leaders base their evaluation and assessment of signals sent by other nations on what they know about particular leaders (Hall & Yarhi-Milo, 2012). Does, for example, reputation attach to institutions, states, governments, publics, and leaders or to all, some, or none of the above? Hall and Yarhi-Milo’s work finds that evaluations of the sincerity of signals are not just based on whether that signal is costly, but also on the evaluator’s personal impressions of the leaders. Reagan’s famous decision to “trust but verify” the claims of Soviet leaders in the context of the Cold War was in part shaped by his personal experiences with Gorbachev. While not much has been done to explore signaling in the individual context, this work offers a useful effort to incorporate neuroscientific research on effective communication into international relations studies more generally.

One of the new frontiers that can bridge the gap between levels of analysis in signaling in world politics aims to disaggregate the “message” portion of the sender–receiver–message framework for signaling. The goals of classical deterrence theory are at least two-fold: to avoid war and to maintain the status-quo division of resources between states. Leaders vary in their objectives and therefore must also vary in their use of signaling designed to achieve these goals. As discussed, winning a dispute (if it happens) and warning an adversary away from using force (to avoid an unnecessary dispute) are two distinct outcomes that depend on different types of leader calculations. Analyzing how actors choose to signal their objectives can provide major insights as to how motives and goals are exercised, how actors signal and under what circumstances, and how signals differ in the degree to which they accomplish an intended goal. These findings could be generalized to characterize certain types of individuals, institutions, interests, or nation-states.

This more extensive approach to analysis of actors and interests has been made possible thanks to recent advances in both theory and data. Researchers have begun to consider heterogeneous ends and means in signaling. Similarly, recent data projects have begun disaggregating aspects of both nuclear and conventional force structures. Such work promises to refine the understanding of concepts such as costly signals, tying hands, and sinking costs. Investigating variation in the means of signaling can lead to better comprehension of where concepts differ in particular circumstances and can also develop new theories about the role of signaling in international interactions.

Looking Ahead and Future Challenges

Much of the early literature on signaling was methodologically bifurcated and homogeneous within those bifurcations. Research either took the form of formal models sprinkled with an occasional illustrative case study (Fearon, 1995; Powell, 2006; Wagner, 2000; Slantchev, 2011) or employed qualitative methods such as case studies or process-tracing in assessing signaling during historical events such as covert interventions by the United States and the Soviet Union in Angola and Afghanistan (Carson & Yarhi-Milo, 2017), the Vietnam War (Sagan & Suri, 2003), or the Korean War (Sartori, 2002). Despite natural complementarities across these approaches, the authors of formal models rarely tested the implications of their models using careful qualitative or quantitative techniques. Similarly, authors who have analyzed specific case studies are often reluctant to speak to the broader validity of their findings for theories about signaling in foreign policy.

Part of the explanation may involve practical features of academic culture. The amount of space needed to explain formal models leaves little room for equally developed empirical work. If that is the case, then new research on signaling should seek to test these models. For example, Fearon and Wagner's different conceptions of war as either integral to or an alternative to bargaining can be tested empirically by looking at how negotiated settlement offers differ prior to war and at the end of a contest. How countries thought about "dividing the pie" when both actors had incentives to misrepresent their resolve, intentions, and/or capabilities versus how they thought about that division after that information was revealed through costly conflict could shed light on issues of reputation, especially if examined through the lens of state expectations of future war with other actors. Similarly, Slantchev's (2012) claim that governments have an incentive to borrow money to coerce opponents into larger concessions should be tested empirically. Here, signaling plays an implied role because the act of financing war via borrowing credibly demonstrates a willingness to depart from what would otherwise have been a mutually acceptable peace, making war more likely. This idea can be empirically tested with data about how states have financed particular wars and how this process changes over time, by dyad, and by the "pie" over which they fight.

Previous research on signaling in foreign policy has set an important foundation on which future research will build. Understanding signaling as occurring within a structure of sender, receiver, message, and reaction as well as organizing concepts like tying hands, sinking costs, cheap talk, and audience costs structure inquiry and determine the form and effectiveness of subsequent scholarship. Researchers are beginning to relax other relationships (bilateral and instrumental) to explore variation in the means and outcomes of signaling. Fortunately, improvements in data are also providing exciting opportunities for advancing scholarship. Access to evidence-based inference will allow the development of a "virtuous cycle" between theory and empirics that has been the hallmark of

cumulative research and important advances in the social sciences. Benefits from, and participation in, the evolution of this important literature are clearly on the horizon.

Further Reading

Clare, J., & Danilovic, V. (2010). **Multiple audiences and reputation building in international conflicts**. *Journal of Conflict Resolution*, *54*(6), 860–882.

Crawford, V. P., & Sobel, J. (1982). **Strategic information transmission**. *Econometrica*, *50*(6), 1431–1451.

Levy, J. S. (2012). **Coercive threats, audience costs, and case studies**. *Security Studies*, *21*(3), 383–390.

Licari, M. J., & Meier, K. J. (2000). **Regulation and signaling: When a tax is not just a tax**. *Journal of Politics*, *62*(3), 875–885.

McManus, R. W. (2016). **The impact of context on the ability of leaders to signal resolve**. *International Interactions*, *43*(3), 453–479.

Miller, G. D. (2011). *The shadow of the past: Reputation and military alliances before the First World War*. Ithaca, NY: Cornell University Press.

O’Neill, B. (1991). Conflictual moves in bargaining: Warnings, threats, escalations, and ultimatums. In H. P. Young (Ed.), *Negotiation analysis* (pp. 87–108). Ann Arbor: University of Michigan Press.

Quek, K. (2013). **Are costly signals more credible? Evidence from three experiments**. *SSRN Electronic Journal*.

Quek, K. (2015). **Discontinuities in signaling behavior upon the decision for war: An analysis of China’s prewar signaling behavior**. *International Relations of the Asia-Pacific*, *15*(2), 279–317.

Quek, K. (2016). **Are costly signals more credible? Evidence of sender-receiver gaps**. *Journal of Politics*, *78*(3), 925–940.

Trager, R. F. (2013). **How the scope of a demand conveys resolve**. *International Theory*, *5*(3), 414–445.

Weeks, J. L. (2008). Autocratic audience costs: Regime type and signaling resolve. *International Organization*, *62*(1), 35–64.

Whang, T. (2010). **Empirical implications of signaling models: Estimation of belief updating in international crisis bargaining**. *Political Analysis*, *18*(3), 381–402.

Wolford, S. (2014). **Showing restraint, signaling resolve: Coalitions, cooperation, and crisis bargaining**. *American Journal of Political Science*, *58*(1), 144–156.

Yarhi-Milo, K. (2013). **Tying hands behind closed doors: The logic and practice of secret reassurance**. *Security Studies*, *22*(3), 405–435.

References

Bemelmans-Videc, M.-L., Rist, R. C., & Vedung, E. O. (2011). *Carrots, sticks, and sermons: Policy instruments and their evaluation*. Piscataway, NJ: Transaction Publishers.

Blainey, Geoffrey. (1973). *The causes of war*. New York: Free Press.

Carson, A., & Yarhi-Milo, K. (2017). **Covert communication: The intelligibility and credibility of signaling in secret**. *Security Studies*, *26*(1), 124–156.

Cetinyan, R. (2002). *Not seeing is believing: What game theory says about reputation and deterrence*. Pittsburgh, PA: University of Pittsburgh.

Chapman, T. L. (2009). Audience beliefs and international organization legitimacy. *International Organization*, *63*(4), 733–764.

Chiozza, G., & Goemans, H. E. (2011). Leaders and international conflict. Cambridge, U.K.: Cambridge University Press.

Cho, I.-K., & Kreps, D. M. (1987). **Signaling games and stable equilibria**. *Quarterly Journal of Economics*, *102*(2), 179–221.

Chu, Jonathan. (2017). *International organizations and information transmission: Evidence from Japan*. Palo Alto, CA: Stanford University Press.

Crescenzi, Mark J. C. (2007). Reputation and interstate conflict. *American Journal of Political Science*, *51*(2), 382–396.

Crescenzi, Mark J. C., Kathman, J. D., Kleinberg, K. B., Wood, R. M. (2012). Reliability, reputation, and alliance formation. *International Studies Quarterly*, *56*(2), 259–274.

Downs, G., & Rocke, D. M. (1995). *Optimal imperfection: Domestic uncertainty and institutions in international relations*. Princeton, NJ: Princeton University Press.

Fearon, J. D. (1994). Domestic political audiences and the escalation of international disputes. *American Political Science Review*, *88*(3), 577–592.

Fearon, J. D. (1995). **Rationalist explanations for war**. *International Organization*, *49*(3), 379–414.

Fearon, J. D. (1997). **Signaling foreign policy interests tying hands versus sinking costs**. *Journal of Conflict Resolution*, *41*(1), 68–90.

Fuhrmann, M., & Sechser, T. S. (2014). **Signaling alliance commitments: Hand-tying and sunk costs in extended nuclear deterrence**. *American Journal of Political Science*, *58*(4), 919–935.

Gartzke, E., Douglass, R., & Lindsay, J. R. (forthcoming). *The Influence of seapower on politics: Domain- and platform-specific attributes of material capabilities*. San Diego: University of California Press.

Gartzke, E., Kaplow, J. M., & Mehta, R. N. (2014). **The determinants of nuclear force structure**. *Journal of Conflict Resolution*, *58*(3), 481–508.

Gartzke, E., Kaplow J. M., & Mehta, R. N. (2017). *Offense, defense and the structure of nuclear forces: The role of nuclear platform diversification in securing second strike*. San Diego: University of California Press.

Gartzke, E., & Lupu, Y. (2012). **Still looking for audience costs**. *Security Studies*, *21*(3), 391–397.

Gartzke, E., & Westerwinter, O. (2016). The complex structure of commercial peace: Contrasting trade interdependence, asymmetry and multipolarity. *Journal of Peace Research*, *53*(3), 325–343.

George, A. L., & Smoke, R. (1974). *Deterrence in American foreign policy: Theory and practice*. New York: Columbia University Press.

Gibler, Douglas M. (2008). The costs of reneging: Reputation and alliance formation. *Journal of Conflict Resolution*, *52*(3), 426–454.

Guisinger, A., & Smith, A. (2002). **Honest threats: The interaction of reputation and political institutions in international crises**. *Journal of Conflict Resolution*, *46*(2), 175–200.

Hall, P., & Franzese, R. (1998). Mixed signals: Central bank independence, coordinated wage bargaining, and European Monetary Union. *International Organization*, *52*(3), 505–535.

Hall, T., & Yarhi-Milo, K. (2012). **The personal touch: Leaders' impressions, costly signaling, and assessments of sincerity in international affairs**. *International Studies Quarterly*, *56*(3), 560–573.

Hamilton, E. J. (2017). **Democratic domestic institutions and foreign policy**. In Cameron Thies (Ed.), *Oxford encyclopedia of foreign policy analysis*.

Handberg, R. (2016). **Aircraft carrier or ballistic missile defense: Signaling commitment in situations of uncertainty**. *Comparative Strategy*, *35*(5), 355–362.

Horowitz, M. C., & Stam, A. C. (2014). **How prior military experience influences the future militarized behavior of leaders**. *International Organization*, *68*(3), 527–559.

Huth, P., & Russett, B. (1984). **What makes deterrence work? Cases from 1900 to 1980**. *World Politics*, *36*(4), 496–526.

Huth, P., & Russett, B. (1988). **Deterrence failure and crisis escalation**. *International Studies Quarterly*, *32*(1), 29–45.

Huth, P., & Russett, B. (1990). **Testing deterrence theory: Rigor makes a difference**. *World Politics*, *42*(4), 466–501.

Huth, P. K. (1998). *Standing your ground: Territorial disputes and international conflict*. Ann Arbor: University of Michigan Press.

Jervis, R. (1976). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.

Jervis, R. (1978). **Cooperation under the security dilemma**. *World Politics*, *30*(2), 167–214.

Jervis, R. (1989). *The logic of images in international relations*. New York: Columbia University Press.

Keohane, R. O. (1986). **Reciprocity in international relations**. *International Organization*, *40*(1), 1–27.

Kertzer, J. D., & Brutger, R. (2015). **Decomposing audience costs: Bringing the audience back into audience cost theory**. *American Journal of Political Science*, *60*(1), 234–249.

Kurizaki, S. (2007). **Efficient secrecy: Public versus private threats in crisis diplomacy**. *American Political Science Review*, *101*(3), 543–558.

Lai, B. (2004). The effects of different types of military mobilization on the outcome of international crises. *Journal of Conflict Resolution*, *48*(2), 211–229.

Lebow, R. N., & Stein, J. G. (1989). **Rational deterrence theory: I think, therefore I deter**. *World Politics*, *41*(2), 208–224.

Leeds, B. A., & Davis, D. (1997). Domestic political vulnerability and international disputes. *Journal of Conflict Resolution*, *41*(6), 814–834.

Levendusky, M. S., & Horowitz, M. C. (2012). When backing down is the right decision: Partisanship, new information, and audience costs. *Journal of Politics*, *74*(2), 323–338.

Lieberman, E. (1994). **The rational deterrence theory debate: Is the dependent variable elusive?***Security Studies*, *3*(3), 384–427.

Lieberman, E. (1995). *Deterrence theory: Success or failure in Arab-Israeli wars*. N.p.: Diane Publishing.

Lindsay, J. R. & Gartzke, E. (forthcoming). Cross-domain deterrence as a practical problem and a theoretical concept. In E. Gartzke and J. R. Lindsay (Eds.), *Cross-domain deterrence: Strategy in an era of complexity*. San Diego: University of California Press.

McGillivray, F., & Smith, A. (2000). **Trust and cooperation through agent-specific punishments**. *International Organization*, *54*(4), 809–824.

Mearsheimer, J. J. (1983). *Conventional deterrence*. Ithaca, NY: Cornell University Press.

Mercer, J. (1996). *Reputation and international politics*. Ithaca, NY: Cornell University Press.

Mercer, J. (1997). **Reputation and rational deterrence theory**. *Security Studies*, *7*(1), 100–113.

Miller, R. A. (2016). **Audience costs, news media, and foreign policy**. In *Oxford encyclopedia of foreign policy analysis*.

Milner, H. V. (1997). *Interests, institutions, and information: Domestic politics and international relations*. Princeton, NJ: Princeton University Press.

Morrow, J. D. (1999). The strategic setting of choices: signaling, commitment, and negotiation in international politics. In D. Lake and R. Powell (Ed.), *Strategic choice and international relations* (pp. 77–114). Princeton, NJ: Princeton University Press.

Orme, J. D. (1992). *Deterrence, reputation and Cold-War cycles*. New York: Springer.

Partell, P. J., & Palmer, G. (1999). Audience costs and interstate crises: An empirical assessment. *International Studies Quarterly*, *43*(2), 389–405.

Powell, R. (1990). *Nuclear deterrence theory: The search for credibility*. Cambridge, U.K.: Cambridge University Press.

Powell, R. (2006). War as a commitment problem. *International Organization*, *60*(1), 169–203.

Press, Daryl Grayson. *Calculating credibility: How leaders assess military threats*. Ithaca, NY: Cornell University Press, 2005.

Ramsay, K. W. (2004). Politics at the water’s edge: Crisis bargaining and electoral competition. *Journal of Conflict Resolution*, *48*(4), 459–486.

Ramsay, K. W. (2017). Negotiation in foreign policy. In C. Thies (Ed.), *Oxford encyclopedia of foreign policy analysis*.

Sagan, S. D., & Suri, J. (2003). The madman nuclear alert: Secrecy, signaling, and safety in October 1969. *International Security*, *27*(4), 150–183.

Sartori, A. E. (2002). The might of the pen: A reputational theory of communication in international disputes. *International Organization*, *56*(1), 121–149.

Sartori, A. E. (2005). *Deterrence by diplomacy*. Princeton, NJ: Princeton University Press.

Saunders, E. N. (2009). **Transformative choices: Leaders and the origins of intervention strategy**. *International Security*, *34*(2), 119–161.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Schelling, T. C. (1966). *Arms and influence*. New Haven, CT: Yale University Press.

Schultz, K. A. (1998). **Domestic opposition and signaling in international crises**. *American Political Science Review*, *92*(4), 829–844.

Schultz, K. A. (2001a). Democracy and coercive diplomacy. Cambridge, MA: Cambridge University Press.

Schultz, K. A. (2001b). **Looking for audience costs**. *Journal of Conflict Resolution*, *45*(1), 32–60.

Shimshoni, J. (1988). *Israel and conventional deterrence: Border warfare from 1953 to 1970*. Ithaca, NY: Cornell University Press.

Slantchev, B. L. (2005). **Military coercion in interstate crises**. *American Political Science Review*, *99*(4), 533–547.

Slantchev, B. L. (2006). **Politicians, the media, and domestic audience costs**. *International Studies Quarterly*, *50*(2), 445–477.

Slantchev, B. L. (2010). **Feigning weakness**. *International Organization*, *64*(3), 357–388.

Slantchev, B. L. (2011). *Military threats: The costs of coercion and the price of peace*. Cambridge, U.K.: Cambridge University Press.

Slantchev, B. L. (2012). **Borrowed power: Debt finance and the resort to arms**. *American Political Science Review*, *106*(4), 787–809.

Smith, A. (1998). **International crises and domestic politics**. *American Political Science Review*, *92*(3), 623–638.

Snyder, J., & Borghard, E. D. (2011). **The cost of empty threats: A penny, not a pound**. *American Political Science Review*, *105*(3), 437–456.

Suong, C., Desposato, S., & Gartzke, E. (2017). *How "democratic" is the democratic peace?: A survey experiment of foreign policy preferences in Brazil and China*. San Diego: University of California Press.

Tago, A., & Ikeda, M. (2015). An “A” for effort: Experimental evidence on UN Security Council engagement and support for US military action in Japan. *British Journal of Political Science*, *45*(2), 391–410.

Thompson, A. (2006). Coercion through IOs: The Security Council and the logic of information transmission. *International Organization*, *60*(1), 1–34.

Tingley, D. H., & Walter, B. F. (2011). **The effect of repeated play on reputation building: An experimental approach**. *International Organization*, *65*(2), 343–365.

Tomz, M. (2007). **Domestic audience costs in international relations: An experimental approach**. *International Organization*, *61*(4), 821–840.

Trachtenberg, M. (2012). Audience costs: An historical analysis. *Security Studies*, *21*(1), 3–42.

Trager, R. F. (2002). *Democratic oppositions have depth: Valence politics in international crises*. Columbia University.

Trager, R. F. (2010). **Diplomatic calculus in anarchy: How communication matters**. *American Political Science Review*, *104*(2), 347–368.

Trager, R. F. (2015). **Diplomatic signaling among multiple states**. *Journal of Politics*, *77*(3), 635–647.

van Belle, D. (1997). Press freedom and the democratic peace. *Journal of Peace Research*, *34*(4), 405–414.

Wagner, R. H. (2000). **Bargaining and war**. *American Journal of Political Science*, *44*(3), 469–484.

Waltz, K. N. (1981). **The spread of nuclear weapons: More may be better: Introduction**. *The Adelphi Papers*, *21*(171), 1.

Weeks, J. L. (2008). Autocratic audience costs: Regime type and signaling resolve. *International Organization*, *62*(1), 65–101.

Weeks, J. L., & Crunkilton, C. (2017). Domestic constraints on foreign policy in authoritarian systems. In C. Thies (Ed.), *Oxford encyclopedia of foreign policy analysis*.

Weiss, J. C. (2013). Authoritarian signaling, mass audiences and national protest in China. *International Organization*, *67*(1), 1–35.

Notes:

(1.) Alternatively, McGillivray and Smith (2000) suggest a mechanism connecting international to domestic audience costs. States may punish dishonest leaders by refusing

to work with their state until the leader is replaced. This action would increase the incentive for domestic audiences to replace the leader and therefore increase domestic audience costs through international reputation pressures.

(2.) Similar conclusions were drawn by Shimshoni, 1988; Orme, 1992; and Lieberman, 1994, 1995.

Erik A. Gartzke

Department of Political Science, University of California, San Diego

Shannon Carcelli

Department of Political Science, University of California, San Diego

J Andres Gannon

Department of Political Science, University of California, San Diego

Jiakun Jack Zhang

Department of Political Science, University of California, San Diego